

STIC-Biotech/ChemLib

69910

Fr m:
Sent:
To:
Subject:

Nguyen, Quang (AU1632)
Monday, June 03, 2002 3:14 PM
STIC-Biotech/ChemLib
Sequence search request for 09/721543

Good afternoon,

Please search:

SEQ ID NO:12, 13, 14, 15 and 16 against commercial, issued and pending U.S. application databases.
Should there be questions, please contact me at 308-8339 (CM1-12A12, AU 1636, my mail box is in CM1-11E12).

THANK YOU.

RECEIVED
JUN - 3 2002
(STIC)

Point of Contact:
Barb O'Bryen
Technical Information Specialist
STIC CM1 6A05 308-4291

Searcher: Bob
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: 6-5-02
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST(where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 17:09:09 ; Search time 4272.65 Seconds
(without alignments)
116.880 Million cell updates/sec

Title: US-09-721-543A-16
Perfect score: 37
Sequence: 1 ctatatccaccatattcccttcccttggtggtggg 37

Scoring table:
IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	23	62.2	513	12	AZ908796
C 2	22.8	61.6	278	12	AZ742846
C 3	22.8	61.6	626	12	AQ348364
C 4	22.6	61.1	564	10	BI682748
C 5	22.6	61.1	934	12	AQ746618
C 6	22	59.5	359	9	AI381292
C 7	21.8	58.9	197	9	AW531399
C 8	21.8	58.9	281	10	BF416687
C 9	21.8	58.9	306	9	AW433771
C 10	21.8	58.9	369	9	AV656658
C 11	21.8	58.9	370	9	AA997381
C 12	21.8	58.9	478	10	BG381507
C 13	21.8	58.9	504	9	AI575981
C 14	21.8	58.9	696	9	AV752617
C 15	21.8	58.9	916	10	BE965057
C 16	21.8	58.9	947	10	BE911323
C 17	21.8	58.9	1275	10	BF688465

RESULT 1
AZ908796/c
LOCUS
DEFINITION
ACCESSION
AZ908796
VERSION
AZ908796.1 GI:13227741
KEYWORDS
GSS
SOURCE
house mouse.
ORGANISM
Mus musculus
REFERENCE
1 (bases 1 to 513)
AUTHORS
Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M., Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E., Russel,D., de Jong,P. and Fraser,C.M.
TITLE
Mouse BAC End Sequences from library RPCI-24
JOURNAL
Unpublished (1999)
COMMENT
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end page: http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
Plate: 222 row: E column: 16
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1. .513
/organism="Mus musculus"
/strain="CS7BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-222E16"
/clone_lib="RPCI-24"

FEATURES
source

C 18 21.6 58.4 288 9 AW242494
C 19 21.6 58.4 376 10 BF226236
C 20 21.6 58.4 523 9 AII151930
C 21 21.6 58.4 544 10 BF286029
C 22 21.6 58.4 600 10 BI990118
C 23 21.6 58.4 683 9 BB140912
C 24 21.6 58.4 698 10 BG923513
C 25 21.6 58.4 762 10 BG620158
C 26 21.6 58.4 764 10 BI688856
C 27 21.6 58.4 815 10 BI645932
C 28 21.6 58.4 822 10 BI522467
C 29 21.4 57.8 214 12 BH626108
C 30 21.4 57.8 265 9 AW353241
C 31 21.4 57.8 348 9 AW336268
C 32 21.4 57.8 379 12 BH628904
C 33 21.4 57.8 442 9 AW478514
C 34 21.4 57.8 445 10 BE753768
C 35 21.4 57.8 474 12 BH625558
C 36 21.4 57.8 475 12 BH408135
C 37 21.4 57.8 497 10 BF039832
C 38 21.4 57.8 519 10 BM365942
C 39 21.4 57.8 543 9 AW657071
C 40 21.4 57.8 574 12 BH631145
C 41 21.4 57.8 577 10 BI540444
C 42 21.4 57.8 1350 12 AG054991
C 43 21.2 57.3 413 12 AZ274408
C 44 21.2 57.3 420 9 AI328516
C 45 21.2 57.3 480 10 BI490098

ALIGNMENTS

AZ908796 513 bp DNA linear GSS 05-MAR-2001
RPCI-24-222E16.TJ RPCI-24 Mus musculus genomic clone RPCI-24-222E16
, DNA sequence.
AZ908796
AZ908796.1 GI:13227741
GSS
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M., Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E., Russel,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from library RPCI-24
Unpublished (1999)
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end page: http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
Plate: 222 row: E column: 16
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1. .513
/organism="Mus musculus"
/strain="CS7BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-222E16"
/clone_lib="RPCI-24"

/sex="Male"
/cell_type="Spleen/Brain"
/note=vector: pTABAC1; Site_1: BamH1; Site_2: BamH1;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTABAC1 cloning vector at the
BamH1 sites using MboI partially digested male C57BL/6J
DNA."

BASE COUNT 158 a 107 c 112 g 136 t

ORIGIN

Query Match 62.2%; Score 23; DB 12; Length 513;
Best Local Similarity 83.9%; Pred. No. 3.9e+02;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 5 atccaccatcccccttcccccttgcgtgtg 35

|||||

Db 270 AGCCTCCCATATCCCTTCCTTCGACGTGTG 240

RESULT 2

AZ742846/c

LOCUS AZ742846 278 bp DNA linear GSS 25-JAN-2001
DEFINITION RPCI-24-110L21-TV RPCI-24 Mus musculus genomic clone RPCI-24-110L21
DNA sequence.

ACCESSION AZ742846

VERSION AZ742846.1 GI:12522124

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 1 (bases 1 to 278)

AUTHORS Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,

Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,

Russell,D., de Jong,P. and Fraser,C.M.

Mouse BAC End Sequences from Library RPCI-24

Unpublished (1999)

CONTACT: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong

(pdejong@mail.cho.org). Clones may be purchased from BACPAC

Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end

page: http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html

Plate: 110 row: L column: 21

Seq primer: T7

Class: BAC ends.

FEATURES

source

Location/Qualifiers

1..278

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-24-110L21"

/clone_lib="RPCI-24"

/sex="Male"

/cell_type="Spleen/Brain"

/note=vector: pTABAC1; Site_1: BamH1; Site_2: BamH1;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTABAC1 cloning vector at the
BamH1 sites using MboI partially digested male C57BL/6J

DNA."

BASE COUNT 100 a 32 c 80 g 66 t

ORIGIN

Query Match 61.6%; Score 22.8; DB 12; Length 278;
Best Local Similarity 92.3%; Pred. No. 4e+02;

Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 10 cccatcccccttcccccttgcgtgtg 35

|||||

Db 101 CCCATATCCCTTCCTTCCTTCCTGTG 76

RESULT 3

AQ348364/c

LOCUS AQ348364 626 bp DNA linear GSS 07-MAY-1999
DEFINITION RPCI11-137E20-TV RPCI-11 Homo sapiens genomic clone RPCI-11-137E20,
DNA sequence.

ACCESSION AQ348364

VERSION AQ348364.1 GI:4173260

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 626)

AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter

,J.C.

Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
Map Building

Unpublished (1997)

CONTACT: Shaying Zhao, William Nierman, Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850

Tel: 301 838 0200

Fax: 301 838 0208

Email: hbe@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html

Seq primer: T7

Class: BAC ends.

FEATURES

source

Location/Qualifiers

1..626

/organism="Homo sapiens"

/db_xref="GDB:752339"

/db_xref="taxon:9606"

/clone="RPCI-11-137E20"

/clone_lib="RPCI-11"

/sex="Male"

/cell_type="Lymphocytes"

/note=vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;

RPCI11 Human Male BAC Library"

BASE COUNT 188 a 162 c 130 g 144 t

ORIGIN

Query Match 61.6%; Score 22.8; DB 12; Length 636;

Best Local Similarity 79.4%; Pred. No. 4.8e+02;

Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 4 tatccaccatcccccttcccccttgcgtgtg 37

|||||

Db 282 TGTCCTTCCATATTCCTTCTCTTGTGCTGGG 249

RESULT 4

BI682748/c

LOCUS BI682748 564 bp mRNA linear EST 17-SEP-2001

DEFINITION 463908 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.

ACCESSION BI682748

VERSION BI682748.1 GI:15635682

KEYWORDS EST.

SOURCE COW.

ORGANISM Bos taurus

REFERENCE AUTHORS	TITLE	JOURNAL MEDLINE	COMMENT
1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4
5	5	5	5
6	6	6	6
7	7	7	7
8	8	8	8
9	9	9	9
10	10	10	10
11	11	11	11
12	12	12	12
13	13	13	13
14	14	14	14
15	15	15	15
16	16	16	16
17	17	17	17
18	18	18	18
19	19	19	19
20	20	20	20
21	21	21	21
22	22	22	22
23	23	23	23
24	24	24	24
25	25	25	25
26	26	26	26
27	27	27	27
28	28	28	28
29	29	29	29
30	30	30	30
31	31	31	31
32	32	32	32
33	33	33	33
34	34	34	34
35	35	35	35
36	36	36	36
37	37	37	37
38	38	38	38
39	39	39	39
40	40	40	40
41	41	41	41
42	42	42	42
43	43	43	43
44	44	44	44
45	45	45	45
46	46	46	46
47	47	47	47
48	48	48	48
49	49	49	49
50	50	50	50
51	51	51	51
52	52	52	52
53	53	53	53
54	54	54	54
55	55	55	55
56	56	56	56
57	57	57	57
58	58	58	58
59	59	59	59
60	60	60	60
61	61	61	61
62	62	62	62
63	63	63	63
64	64	64	64
65	65	65	65
66	66	66	66
67	67	67	67
68	68	68	68
69	69	69	69
70	70	70	70
71	71	71	71
72	72	72	72
73	73	73	73
74	74	74	74
75	75	75	75
76	76	76	76
77	77	77	77
78	78	78	78
79	79	79	79
80	80	80	80
81	81	81	81
82	82	82	82
83	83	83	83
84	84	84	84
85	85	85	85
86	86	86	86
87	87	87	87
88	88	88	88
89	89	89	89
90	90	90	90
91	91	91	91
92	92	92	92
93	93	93	93
94	94	94	94
95	95	95	95
96	96	96	96
97	97	97	97
98	98	98	98
99	99	99	99
100	100	100	100

COMMENT	FEATURES	SOURCE
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 102 a 80 c 108 g 67 t 58.9%; Score 22; DB 9; Length 359; Best Local Similarity 78.1%; Pred. No. 7.9e+02; Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 6 tccaccatccccctcccttcggtggtg 37
 Db 314 TCCACCATGCCCATCCCTTCATGTGNG 345

RESULT 7
 LOCUS AW531399/c 197 bp mRNA linear EST 06-MAR-2000
 DEFINITION UI-R-BS0-amm-e-05-0-UI.s1 UI-R-BS0 Rattus norvegicus cDNA clone

ACCESSION AW531399
 VERSION AW531399.1 GI:7173813
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 197)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the non-normalized embryo at 13 dpc library cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA=Yes.

FEATURES
 source Location/Qualifiers
 1..197
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-BS0-amm-e-05-0-UI"
 /clone_lib="UI-R-BS0"
 /dev_stage="embryonic 13 dpc"
 /lab_host="DH10B (Life Technologies)"
 /notes="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BS0 library is derived from 13 dpc whole embryo tissue. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu."
 TAG_LIB=UI-R-BS0
 TAG_TISSUE=embryo at 13 dpc
 TAG_SEQ=AATCC

BASE COUNT 47 a 35 c 62 g 53 t
 ORIGIN

Query Match 58.9%; Score 21.8; DB 9; Length 197; Best Local Similarity 78.8%; Pred. No. 8.1e+02; Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ctatatccaccatcccccttccttcggtg 33
 Db 112 CTATCTCCCCCATGTCCCATGCCCTTCCCTG 80

RESULT 8
 LOCUS BF416687/c 281 bp mRNA linear EST 28-NOV-2000
 DEFINITION UI-R-CN0-bkz-b-11-0-UI.s1 UI-R-CN0 Rattus norvegicus cDNA clone

ACCESSION BF416687
 VERSION BF416687.1 GI:11404716
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 281)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized fundus library cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com) The following repetitive elements were found in this cDNA sequence: 1-21, >AT-rich#Low_complexity
 Seq primer: M13 Forward
 POLYA=Yes.

FEATURES
 source Location/Qualifiers
 1..281
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-CN0-bkz-b-11-0-UI"
 /clone_lib="UI-R-CN0"
 /dev_stage="ADULT"
 /lab_host="DH10B (Life Technologies)"
 /notes="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CN0 library is a normalized library constructed from the following tissues: brown adipose, penis, salivary glands, bladder, fundus, cervix, seminal vesicles. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
 TAG_LIB=UI-R-CN0
 TAG_TISSUE=fundus
 TAG_SEQ=TTCCG

BASE COUNT 67 a 52 c 86 g 76 t
 ORIGIN

Query Match 58.9%; Score 21.8; DB 10; Length 281;

Best Local Similarity 78.83; Pred. No. 9e+02;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ctatatccaccatatacccccttccccttg:ggtg 33
||||| ||||| ||||| ||||| | ||||| |||

Dp 112 CTATCTCCCCCATGTCCCCCATGCCCTTCCCTG 80

RESULT	10
AV656658/C	

AV556656/C	LOCUS	AV556658	363 bp	mrna	linear	EST 16-JAN-2002
DEFINITION	AV556658	GLC Homo sapiens cDNA clone GLCET05 3', mRNA sequence.				

VERSION AV656658.1
KEYWORDS EST.
SOURCE human

SOURCE
ORGANISM

Eukaryota: M

REFERENCE

1 (bases 1 to 369)

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS

Xiao, H., Qu, J., Liu, F., Huang, Q., Cheng, Z., Li, N., Du, J., Hu, W., Shen, K., Lu, G., Fu, G., Zhong, M., Xu, S., Gu, W., Huang, W., Zhao, X.

TITLE Insight into hepatocellular carcinogenesis at transcriptome level

by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver

JOURNAL	PROC.	Na
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

BOOKNAME MEDLINE
FLOC: NUCL: ACAD: 30
21625106

IDENTIFY
COMMENT

COMMENT
CONTACT: zeyuanyu.han@genomics.cn
Chinese National Human Genome Center at Shanghai

351 GUO Shouting BO

201203 P. R. China

201203, F: K: CHINA
Tel: 86-21-50801919

TEL: 80-21-50801919(ex.43)
FAX: 86-21-50801922

FEATURES
 source
 1. .369
 Location/Qualifiers
 This clone is available at CHGC in Shanghai.
 Email: hanzg@chgc.sh.cn
 Fax: 86-21-50001922
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="GLCETf05"

POLYA=Yes.

/organism="Rattus norvegicus"

/strain="Soragie-Dawley"

/db xref="taxon:10116"

```
/usr_xref= caxon:10110
/clone="UT-B-B.T0n-afq-a-08-0-UT"
```

```
/clone-01 K B0Vp
/clone lib=aut-r-r
```

```
/dev staco="adulst"
```

```
/dev_stage=adult
/lab_host="nu10p.rifs.mechol.ca"

```

Query Match	58.9%;	Score 21.8; DB 9;	Length 369;
Best Local Similarity	78.8%;	Pred. No. 9.4e+02;	
Matches	26; Conservative	0; Mismatches	7; Indels
			0; Gaps

TAG_LIB=UI-R-BJ0p

TAG_TISSUE-AV canal at 16.5 dpc

TAG SEO=GAACC"

BASE COUNT	73 a	56 c	97 a	80 b
------------	------	------	------	------

ORIGIN

2440

Query Match

Query Match 38.5%; Score 41.8; DB 5; Length 306;


```
/clone="IMAGE:3886043"  
/clone_lib="Nih_MGC_89"  
/tissue_type="large cell carcinoma, undifferentiated"  
/lab_host="DH10B (phage-resistant)"  
/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;  
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.1 kb. Library constructed by Life  
Technologies."
```

```
BASE COUNT      321 a      167 c      279 g      149 t  
ORIGIN
```

```
Query Match      58.9%; Score 21.8; DB 10; Length 916;  
Best Local Similarity 78.8%; Pred. NO. 1.1e+03;  
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
```

```
Qy      5 atccaccatattcccttcccttgcgtgtggg 37  
      ||||| ||| | ||||| ||||| ||||| |||||  
Db      670 ATCCGCCCTTCTCCCCCCCCCTTTGGGTGGG 538
```

```
Search completed: June 4, 2002, 17:09:16  
Job time: 4549 sec
```


Qy 5 atccaccatattcccttcccccttgcgtgtggg 37
||||| | | | | | | | | | | | | | | |
Db 339 ATCCACTGTGATGCACCTTCCCTTGGGGTTGG 307

RESULT 2
US-08-481-814A-2/C
; Sequence 2, Application US/08481814A
; Patent No. 5869040
; GENERAL INFORMATION:
; APPLICANT: Hsu, Yen-Ming
; TITLE OF INVENTION: GENE THERAPY METHODS AND COMPOSITIONS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Biogen, Inc.
; STREET: 14 Cambridge Center
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 02142
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/481.814A
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Kaplan, Warren A
; REFERENCE/DOCKET NUMBER: A001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-679-2000
; TELEFAX: 617-679-2838
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1766 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; CELL LINE: HeLa
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 429..1739
; OTHER INFORMATION: /product= "E2F-2"
US-08-481-814A-2

Query Match 54.68; Score 20.2; DB 2; Length 1766;
Best Local Similarity 75.88; Pred. No. 17;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 5 atccaccatattcccttcccccttgcgtgtggg 37
||||| | | | | | | | | | | | | | | |
Db 767 ATCCACTGTGATGCACCTTCCCTTGGGGTTGG 735

RESULT 3
US-08-136-119-3/c
; Sequence 3, Application US/08136119
; Patent No. 5473056
; GENERAL INFORMATION:
; APPLICANT: Heimbrosk, David C.
; APPLICANT: Hoyle, Mona I.
; APPLICANT: Oliff, Allen I.
; TITLE OF INVENTION: E 2F-2, A NOVEL MAMMALIAN TRANSCRIPTION
; TITLE OF INVENTION: FACTOR

; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Muthard
; STREET: P.O. Box 2000, 126 Lincoln Avenue
; CITY: Rahway
; STATE: N.J.
; COUNTRY: USA
; ZIP: 07065-0907
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/136.119
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Muthard, David A.
; REGISTRATION NUMBER: 35,297
; REFERENCE/DOCKET NUMBER: 19042
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908)594-3903
; TELEFAX: (908)594-4720
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2647 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-136-119-3

Query Match 54.68; Score 20.2; DB 1; Length 2647;
Best Local Similarity 75.88; Pred. No. 18;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 5 atccaccatattcccttcccccttgcgtgtggg 37
||||| | | | | | | | | | | | | | | |
Db 767 ATCCACTGTGATGCACCTTCCCTTGGGGTTGG 735

RESULT 4
US-08-845-998-3
; Sequence 3, Application US/08845998
; Patent No. 5879892
; GENERAL INFORMATION:
; APPLICANT: Van Baren, Nicolas
; APPLICANT: Coullie, Pierre G.
; APPLICANT: De Smet, Charles
; APPLICANT: Lucas, Sophie
; APPLICANT: Boon, Thierry
; TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/845.998
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Van Amsterdam, John R.

REGISTRATION NUMBER: 40,212
REFERENCE/DOCKET NUMBER: L0461/7008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)720-3500
TELEFAX: (617)720-2441
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2259 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 8..577
US-08-845-998-3

Query Match 52.4%; Score 19.4; DB 2; Length 2259;
Best Local Similarity 79.3%; Pred. No. 36;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 7 ccaccatattccctcccttgcgtg 35
||||| | ||||| ||| | | |
Db 2139 CCACCTGAGCCCTTCCCTTTCTTGAG 2167

RESULT 5
US-09-206-537-3
Sequence 3, Application US/09206537
Patent No. 6130052
GENERAL INFORMATION:
APPLICANT: Van Baren, Nicolas
APPLICANT: Coulie, Pierre G.
APPLICANT: De Smet, Charles
APPLICANT: Lucas, Sophie
APPLICANT: Boon, Thierry
TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
STREET: 600 Atlantic Avenue
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/206,537
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/845,998
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Van Amsterdam, John R.
REGISTRATION NUMBER: 40,212
REFERENCE/DOCKET NUMBER: L0461/7008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)720-3500
TELEFAX: (617)720-2441
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2259 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 8..577
US-09-206-537-3

Query Match 52.4%; Score 19.4; DB 3; Length 2259;
Best Local Similarity 79.3%; Pred. No. 36;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 7 ccaccatattccctcccttgcgtg 35
||||| | ||||| ||| | | |
Db 2139 CCACCTGAGCCCTTCCCTTTCTTGAG 2167

RESULT 6
US-09-430-854-3
Sequence 3, Application US/09430854
Patent No. 6271019
GENERAL INFORMATION:
APPLICANT: Van Baren, Nicolas
APPLICANT: Coulie, Pierre G.
APPLICANT: De Smet, Charles
APPLICANT: Lucas, Sophie
APPLICANT: Boon, Thierry
TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
STREET: 600 Atlantic Avenue
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/430,854
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/845,998
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Van Amsterdam, John R.
REGISTRATION NUMBER: 40,212
REFERENCE/DOCKET NUMBER: L0461/7008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)720-3500
TELEFAX: (617)720-2441
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2259 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 8..577
US-09-430-854-3

Query Match 52.4%; Score 19.4; DB 4; Length 2259;
Best Local Similarity 79.3%; Pred. No. 36;

Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 7 ccaccatccccctcccttgctgtg 35
||||| | ||||| ||||| |||
Db 2139 CCACCTGAGCCCTCCCTTCTTGAG 2167

RESULT 7

US-08-756-506-23/c
; Sequence 23, Application US/08756506
; Patent No. 5905185

GENERAL INFORMATION:

; APPLICANT: Garner, Ian
; APPLICANT: Cottingham, Ian R.
; APPLICANT: Temperley, Simon M.
; APPLICANT: Foster, Donald C.
; APPLICANT: Sprecher, Cindy A.
; APPLICANT: Prunkard, Donna E.
; TITLE OF INVENTION: PROTEIN C PRODUCTION IN TRANSGENIC
; TITLE OF INVENTION: ANIMALS
; NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA

; COUNTRY: USA

; ZIP: 98102

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/756,506

; FILING DATE:

; CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

; NAME: Sawislak, Deborah A

; REGISTRATION NUMBER: 37,438

; REFERENCE/DOCKET NUMBER: 95-28

TELECOMMUNICATION INFORMATION:

; TELEPHONE: 206-442-6672

; TELEFAX: 206-442-6678

; INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:

; LENGTH: 6763 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

US-08-756-506-23

Query Match 52.4%; Score 19.4; DB 2; Length 6763;
Best Local Similarity 79.3%; Pred. No. 43;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 7 ccaccatccccctcccttgctgtg 35
||||| | ||||| ||||| |||||
Db 4342 CCACCAAGACCCCTACCCCATGTGTG 4314

RESULT 8

US-08-206-176-7/c

; Sequence 7, Application US/08206176

; Patent No. 5639940

GENERAL INFORMATION:

; APPLICANT: Garner, Ian

; APPLICANT: Dairymple, Michael A

; APPLICANT: Prunkard, Donna E

; APPLICANT: Foster, Donald C

; TITLE OF INVENTION: Production of Fibrinogen in Transgenic

; TITLE OF INVENTION: Animals

; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 4225 Roosevelt Way, N.E.
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/206,176
; FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

; NAME: Parker, Gary E

; REGISTRATION NUMBER: 31-648

; REFERENCE/DOCKET NUMBER: 93-15

TELECOMMUNICATION INFORMATION:

; TELEPHONE: 206-547-8080 ext 322

; TELEFAX: 206-548-2329

; INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

; LENGTH: 10807 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; CLONE: ovine beta-lactoglobulin

US-08-206-176-7

Query Match 52.4%; Score 19.4; DB 1; Length 10807;
Best Local Similarity 79.3%; Pred. No. 46;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 7 ccaccatccccctcccttgctgtg 35
||||| | ||||| ||||| |||||
Db 8386 CCACCAAGACCCCTACCCCATGTGTG 8358

RESULT 9

US-08-756-506-5/c

; Sequence 5, Application US/08756506

; Patent No. 5905185

GENERAL INFORMATION:

; APPLICANT: Garner, Ian

; APPLICANT: Cottingham, Ian R.

; APPLICANT: Temperley, Simon M.

; APPLICANT: Foster, Donald C.

; APPLICANT: Sprecher, Cindy A.

; APPLICANT: Prunkard, Donna E.

; TITLE OF INVENTION: PROTEIN C PRODUCTION IN TRANSGENIC

; TITLE OF INVENTION: ANIMALS

; NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

; ADDRESSEE: ZymoGenetics, Inc.

; STREET: 1201 Eastlake Avenue East

; CITY: Seattle

; STATE: WA

; COUNTRY: USA

; ZIP: 98102

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/756,506

; FILING DATE:

CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Sawislak, Deborah A
REGISTRATION NUMBER: 37,438
REFERENCE/DOCKET NUMBER: 95-28
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-442-6672
TELEFAX: 206-442-6678
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 10807 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-756-506-5

Query Match 52.4%; Score 19.4; DB 2; Length 10807;
Best Local Similarity 79.3%; Pred. No. 46;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 7 ccaccatccccctcccttcgctgtg 35
||||| ||||| ||||| |||||
Db 8386 CCACCAAGACCCCTACCCCATGTGTG 8358

RESULT 10
US-09-613-182-13/c
; Sequence 13, Application US/09613182
; Patent No. 6294653
; GENERAL INFORMATION:
; APPLICANT: Mayfield, Stephen
; TITLE OF INVENTION: RNA BINDING PROTEIN AND BINDING SITE USEFUL FOR
; FILE REFERENCE: SCR2177S
; CURRENT APPLICATION NUMBER: US/09/613,182
; CURRENT FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: 09/341,550
; PRIOR FILING DATE: 1999-07-13
; PRIOR APPLICATION NUMBER: PCT/US98/00840
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 60/035,955
; PRIOR FILING DATE: 1997-01-17
; PRIOR APPLICATION NUMBER: 60/069,400
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 1424
; TYPE: DNA
; ORGANISM: Chlamydomonas reinhardtii
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (252)..(1310)
; NAME/KEY: misc_feature
; LOCATION: (279)
; OTHER INFORMATION: Codon also can encode Ser
; NAME/KEY: misc_feature
; LOCATION: (282)
; OTHER INFORMATION: Codon also can encode Glu
; NAME/KEY: misc_feature
; LOCATION: (294)
; OTHER INFORMATION: Codon also can encode Gly
; NAME/KEY: misc_feature
; LOCATION: (306)
; OTHER INFORMATION: Codon also can encode Asn
; NAME/KEY: misc_feature
; LOCATION: (357)
; OTHER INFORMATION: Codon also can encode Leu
; NAME/KEY: misc_feature
; LOCATION: (369)
; OTHER INFORMATION: Codon also can encode Thr
; NAME/KEY: misc_feature

; LOCATION: (486)
; OTHER INFORMATION: Codon also can encode Ser
; NAME/KEY: misc_feature
; LOCATION: (495)
; OTHER INFORMATION: Codon also can encode Ile
; NAME/KEY: misc_feature
; LOCATION: (510)
; OTHER INFORMATION: Codon also can encode Ala
; NAME/KEY: misc_feature
; LOCATION: (555)
; OTHER INFORMATION: Codon also can encode Val
; NAME/KEY: misc_feature
; LOCATION: (588)
; OTHER INFORMATION: Codon also can encode Glu
; NAME/KEY: misc_feature
; LOCATION: (600)
; OTHER INFORMATION: Codon also can encode Leu
; NAME/KEY: misc_feature
; LOCATION: (621)
; OTHER INFORMATION: Codon also can encode Ala
; NAME/KEY: misc_feature
; LOCATION: (714)
; OTHER INFORMATION: Codon also can encode Thr
; NAME/KEY: misc_feature
; LOCATION: (729)
; OTHER INFORMATION: Codon also can encode Ile
; NAME/KEY: misc_feature
; LOCATION: (1092)
; OTHER INFORMATION: Codon also can encode Val
; NAME/KEY: misc_feature
; LOCATION: (1191)
; OTHER INFORMATION: Codon also can encode Ile
; NAME/KEY: misc_feature
; LOCATION: (1284)
; OTHER INFORMATION: Codon also can encode Ala
; NAME/KEY: misc_feature
; LOCATION: (1287)
; OTHER INFORMATION: Codon also can encode Ile
; NAME/KEY: misc_feature
; LOCATION: (1290)
; OTHER INFORMATION: Codon also can encode Glu
; NAME/KEY: misc_feature
; LOCATION: (1293)
; OTHER INFORMATION: Codon also can encode Ala
; NAME/KEY: misc_feature
; LOCATION: (1296)
; OTHER INFORMATION: Codon also can encode Pro
; NAME/KEY: misc_feature
; LOCATION: (1302)
; OTHER INFORMATION: Codon also can encode Thr
; NAME/KEY: misc_feature
; LOCATION: (1308)
; OTHER INFORMATION: Codon also can encode Gly
US-09-613-182-13

Query Match 51.9%; Score 19.2; DB 4; Length 1424;
Best Local Similarity 75.0%; Pred. No. 40;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 tatatccaccatcccccttccttcgctg 33
| ||| ||| | ||||| ||||| |||
Db 117 TTTATGTACCTACGTCCCTTCCCTTACGGG 86

RESULT 11
US-08-466-603-4
; Sequence 4, Application US/08466603
; Patent No. 5726018
; GENERAL INFORMATION:
; APPLICANT: Pasternack, Gary R.
; APPLICANT: Kuhnajda, Francis P.
; TITLE OF INVENTION: No. 5726018el Mammalian Protein Associated With

;; TITLE OF INVENTION: Uncontrolled Cell Division
;; NUMBER OF SEQUENCES: 9
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Banner, Birch, McKie & Beckett
;; STREET: 1001 G Street, N.W.
;; CITY: Washington, D.C.
;; STATE: District of Columbia
;; COUNTRY: U.S.A.
;; ZIP: 20001
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/466,603
;; FILING DATE: 06-JUN-1995
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/314,503
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Posorske Esq., Laurence H.
;; REGISTRATION NUMBER: 34,698
;; REFERENCE/DOCKET NUMBER: 1107.47218
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202 508-9153
;; TELEFAX: 202 508-9299
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 759 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; ORIGINAL SOURCE:
;; ORGANISM: Mus sp
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 3..548
US-08-466-603-4

Query Match 51.4%; Score 19; DB 1; Length 759;
Best Local Similarity 81.5%; Pred. No. 44;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ctatatccaccatccctccctccct 27
||| | ||||| ||||| |||||
Db 591 CTGTTTACCCTATCCCTCCCT 617

RESULT 12
US-08-314-503A-4
;; Sequence 4, Application US/08314503A
;; Patent No. 5734022
;; GENERAL INFORMATION:
;; APPLICANT: Pasternack, Gary R.
;; APPLICANT: Kuhajda, Francis P.
;; TITLE OF INVENTION: No. 5734022el Mammalian Protein Associated With
;; TITLE OF INVENTION: Uncontrolled Cell Division
;; NUMBER OF SEQUENCES: 9
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Banner, Birch, McKie & Beckett
;; STREET: 1001 G Street, N.W.
;; CITY: Washington, D.C.
;; STATE: District of Columbia
;; COUNTRY: U.S.A.
;; ZIP: 20001
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/314,503A
;; FILING DATE:
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Posorske Esq., Laurence H.
;; REGISTRATION NUMBER: 34,698
;; REFERENCE/DOCKET NUMBER: 1107.47218
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202 508-9153
;; TELEFAX: 202 508-9299
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 759 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; ORIGINAL SOURCE:
;; ORGANISM: Mus sp
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 3..548
US-08-314-503A-4

Query Match 51.4%; Score 19; DB 1; Length 759;
Best Local Similarity 81.5%; Pred. No. 44;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ctatatccaccatccctccctccct 27
||| | ||||| ||||| |||||
Db 591 CTGTTTACCCTATCCCTCCCT 617

RESULT 13
US-08-468-066-4
;; Sequence 4, Application US/08468066
;; Patent No. 5756676
;; GENERAL INFORMATION:
;; APPLICANT: Pasternack, Gary R.
;; APPLICANT: Kuhajda, Francis P.
;; TITLE OF INVENTION: No. 5756676el Mammalian Protein Associated With
;; TITLE OF INVENTION: Uncontrolled Cell Division
;; NUMBER OF SEQUENCES: 9
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Banner, Birch, McKie & Beckett
;; STREET: 1001 G Street, N.W.
;; CITY: Washington, D.C.
;; STATE: District of Columbia
;; COUNTRY: U.S.A.
;; ZIP: 20001
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/468,066
;; FILING DATE: 06-JUN-1995
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/314,503
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Posorske Esq., Laurence H.
;; REGISTRATION NUMBER: 34,698
;; REFERENCE/DOCKET NUMBER: 1107.47218
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202 508-9153
;; TELEFAX: 202 508-9299
;; INFORMATION FOR SEQ ID NO: 4:

```

: ATTORNEY/AGENT INFORMATION:
: NAME: PORSORE, Esq., Laurence H.
: REGISTRATION NUMBER: 34,698
: REFERENCE/DOCKET NUMBER: 1107, 47218
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202 508-9153
: TELEFAX: 202 508-9299
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 759 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: ORIGINAL SOURCE:
: ORGANISM: Mus sp
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 3..548
: US-08-466-717-4

```

Search completed: June 4, 2002, 17:47:29
Job time: 6602 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 17:56:22 ; Search time 521.5 Seconds
(without alignments)
121.814 Million cell updates/sec

Title: US-09-721-543A-16
Perfect score: 37
Sequence: 1 ctatataccaccatatacccttcccttgcgtgtggtg 37

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*

- 1: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	22.2	60.0	1302	23	AAS69409		DNA encoding novel
2	22.2	60.0	1302	23	AAS93976		DNA encoding novel
3	22.2	60.0	6337	23	AAS94029		DNA encoding novel
4	21	56.8	3132	22	AAH15189		Human cDNA sequenc
c	5	20.8	56.2	619	21	AAF08347	Fusarium venenatum
c	6	20.8	56.2	1462	22	AAK51597	Human polynucleoti
7	20.6	55.7	1452	23	AAS69410		DNA encoding novel
8	20.6	55.7	1860	23	AAS94021		DNA encoding novel
c	9	20.6	1927	23	AAS91163		DNA encoding novel

C	10	20.6	55.7	2285	23	ABL22366	Drosophila melanog
	11	20.6	55.7	5909	23	AAS63432	DNA encoding novel
C	12	20.6	55.7	6012	23	AAS94024	DNA encoding novel
	13	20.6	55.7	7903	23	AAS77690	DNA encoding novel
C	14	20.4	55.1	13094	22	AAK69540	Human immune/haema
C	15	20.2	54.6	2647	16	AAQ86483	Human transcriptio
C	16	20.2	54.6	11494	24	AAAS18559	Even-skipped homeo
C	17	20.2	54.6	11494	24	AAAS18560	Generic sequence e
	18	20	54.1	492	22	AAH29461	Drosophila melanog
	19	20	54.1	26370	23	ABL07082	Drosophila melanog
	20	19.8	53.5	455	22	ABA70980	Human foetal liver
	21	19.8	53.5	455	22	ABA37400	Probe #15866 for g
	22	19.8	53.5	455	22	AAK19256	Human brain expres
	23	19.8	53.5	455	22	AAK45226	Human bone marrow
	24	19.8	53.5	455	22	AAI25198	Probe #15131 for g
	25	19.8	53.5	455	22	AAI51177	Probe #19863 used
	26	19.8	53.5	512	22	ABA72367	Human foetal liver
	27	19.8	53.5	512	22	ABA38177	Probe #16643 for g
	28	19.8	53.5	512	22	AAK20793	Human brain expres
	29	19.8	53.5	512	22	AAK46943	Human bone marrow
	30	19.8	53.5	512	22	AAI25788	Probe #15721 for g
	31	19.8	53.5	512	22	AAI52780	Probe #21466 used
C	32	19.8	53.5	523	22	ABA08360	Human secreted pro
	33	19.8	53.5	555	22	AAH33403	Human colon cancer
	34	19.8	53.5	600	22	ABA58826	Human foetal liver
	35	19.8	53.5	600	22	ABA26298	Probe #6764 for ge
	36	19.8	53.5	600	22	AAK08099	Human brain expres
	37	19.8	53.5	600	22	AAK33975	Human bone marrow
	38	19.8	53.5	600	22	AAI16592	Probe #6525 for ge
	39	19.8	53.5	600	22	AAI35699	Probe #8385 used t
	40	19.8	53.5	840	23	AAAS86137	DNA encoding novel
	41	19.8	53.5	1096	23	AAAS89756	DNA encoding novel
C	42	19.8	53.5	1125	22	ABAL1802	Human nervous syst
	43	19.8	53.5	1497	23	AAAS86138	DNA encoding novel
	44	19.8	53.5	3423	23	AAAS86141	DNA encoding novel
C	45	19.8	53.5	3899	19	AAV63192	CDNA from clone dt

ALIGNMENTS

RESULT 1

AAS69409
ID AAS69409 standard; cdna; 1302 BP.

XX AC AAS69409;

XX DT 13-FEB-2002 (first entry)

XX DE DNA encoding novel human diagnostic: protein #5213.

XX DE Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

XX DR P-PSDB; ABG05222.

XX PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 1; SEQ ID No 5213; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1302 BP; 414 A; 285 C; 304 G; 299 T; 0 other;

Query Match 60.0%; Score 22.2; DB 23; Length 1302;
 Best Local Similarity 77.1%; Pred. No. 17;
 Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
 Qy 3 atataccaccatccccctcccttcgctgtggg 37
 ||||| ||||| ||||| ||||| ||||| |||||
 Db 1171 atataccaccatctctatcccccttctttatcgg 1205

RESULT 2
 AAS93976
 ID AAS93976 standard; cDNA: 1302 BP.
 XX
 AC AAS93976;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #29780.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR P-PSDB; ABG29789.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 1; SEQ ID No 29780; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1302 BP; 414 A; 285 C; 304 G; 299 T; 0 other;

Query Match 60.0%; Score 22.2; DB 23; Length 1302;
 Best Local Similarity 77.1%; Pred. No. 17;
 Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
 Qy 3 atataccaccatccccctcccttcgctgtggg 37
 ||||| ||||| ||||| ||||| ||||| |||||
 Db 1171 atataccaccatctctatcccccttctttatcgg 1205

RESULT 3
 AAS94029
 ID AAS94029 standard; cDNA: 6337 BP.
 XX
 AC AAS94029;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #29833.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR P-PSDB; ABG29842.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess

PT biodiversity -
 PS Claim 1; SEQ ID NO 29833; 103pp; English.
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 6337 BP; 1813 A; 1504 C; 1508 G; 1512 T; 0 other;

Query Match 60.0%; Score 22.2; DB 23; Length 6337;
 Best Local Similarity 77.1%; Pred. No. 21;
 Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 3 atatcaccatcatcccttccttcgtggtggg 37
 ||||| ||||| ||||| ||||| |||||
 Db 1643 atatcaccatcatccttccttccttcgtggtggg 1677

RESULT 4
 ID AAH15189 standard; cDNA; 3132 BP.
 XX
 AC AAH15189;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA sequence SEQ ID NO:13279.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 XX
 PR 27-AUG-1999; 99JP-0300253.
 XX
 PR 11-JAN-2000; 2000JP-0118776.
 XX
 PR 02-MAY-2000; 2000JP-0183767.
 XX
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-318749/34.
 XX
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 PS Claim 8; SEQ ID 13279; 2537pp + CD ROM; English.
 CC The present invention describes primer sets for synthesising 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 3132 BP; 601 A; 943 C; 769 G; 819 T; 0 other;

Query Match 56.8%; Score 21; DB 22; Length 3132;
 Best Local Similarity 82.8%; Pred. No. 57;
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 atccaccatcatcccttccttcgtggtg 33
 ||||| ||||| ||||| ||||| |||||
 Db 1354 atcccccattctccactacccttcgtg 1382

RESULT 5
 ID AAF08347/C standard; cDNA; 619 BP.
 XX
 AC AAF08347;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Fusarium venenatum EST SEQ ID NO:870.
 XX
 KW Multiple gene expression; filamentous fungal cell; EST;
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KW Aspergillus oryzae; Trichoderma reesei; Identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.
 XX
 OS Fusarium venenatum.
 XX
 PN WO200056762-A2.
 XX
 PD 28-SEP-2000.
 XX
 PF 22-MAR-2000; 2000WO-US07781.
 XX
 PR 22-MAR-1999; 99US-0273623.
 XX
 PA (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO NORDISK AS.
 XX
 PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 XX
 XX WPI; 2000-594572/56.
 DR
 XX

CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 5909 BP; 1569 A; 1426 C; 1483 G; 1431 T; 0 other;

Query Match 55.7%; Score 20.6; DB 23; Length 5909;
Best Local Similarity 74.3%; Pred. No. 89;
Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 3 atatcaccaccatcccccttcccttgctggtggg 37
||||| ||||||| ||||||| |||
Db 3331 atatcaccaccatctcattcccttctttatcg 3365

RESULT 12
AAS94024/c
ID AAS94024 standard; cDNA; 6012 BP.
XX
AC AAS94024;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #29828.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG29837.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID NO 29828; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 6012 BP; 1257 A; 1504 C; 1408 G; 1843 T; 0 other;

Query Match 55.7%; Score 20.6; DB 23; Length 6012;
Best Local Similarity 74.3%; Pred. No. 89;
Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 3 atatcaccaccatcccccttcccttgctggtggg 37
||||| ||||||| ||||||| |||
Db 5426 ATATCAACCATCTCTCTCTCTCTCTCTCT 5392

RESULT 13
AAS77690
ID AAS77690 standard; cDNA; 7903 BP.
XX
AC AAS77690;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #13494.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG13503.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID NO 13494; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 7903 BP; 2316 A; 1835 C; 1850 G; 1902 T; 0 other;

Query Match 55.7%; Score 20.6; DB 23; Length 7903;
Best Local Similarity 74.3%; Pred. No. 92;
Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 3 atatacaccatataccctcccttcgtggtgg 37
Db 5477 atatacaccatatacccttcgtggtgg 5511

RESULT 14

AAK69540/c
ID AAK69540 standard; DNA; 13094 BP.

XX
AC AAK69540;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:24352.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0227009.
PR 01-SEP-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 06-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.

PR 08-NOV-2000; 2000US-0245610.
PR 08-NOV-2000; 2000US-0245611.
PR 08-NOV-2000; 2000US-0245613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2001US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
XX Disclosure; SEQ ID NO 24352; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
XX Sequence 13094 BP; 2675 A; 3841 C; 4019 G; 2559 T; 0 other;

Query Match 55.1%; Score 20.4; DB 22; Length 13094;
Best Local Similarity 80.0%; Pred. No. 1.2e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 8 caccatatacccttcccttgcgtggtg 37
Db 5298 CACCCCATCCACCCCTCGAGTGTG 5269
RESULT 15
AAQ86483/c
ID AAQ86483 standard; cDNA; 2647 BP.
XX AAQ86483;
XX 23-NOV-1995 (first entry)
DE Human transcription factor E2F-2 cDNA.
XX E2F-2; transcription; factor; regulation; cell cycle;
KW retinoblastoma; tumour; suppressor; ss.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH CDS 429..1683
FT /*tag= a
FT /product= E2F-2 transcription factor.
XX
XX GB2282814-A.
XX 19-APR-1995.
XX 07-OCT-1994; 94GB-0020283.
XX 13-OCT-1993; 93US-0136119.
XX (MERI) MERCK & CO INC.
XX Heimbrook DC, Ivey-Hoyle M, Oliff AI;
DR WPI; 1995-141220/19.
DR P-PSDB; AAR71734.
XX
XX New human transcription factor E2F-2 - involved in cell cycle
PT regulation and useful for drug screening, also related cDNA,
PT plasmids and transformed cells.
XX
XX Claim 3; Page 30-31; 53pp; English.
XX
XX AAQ86483 codes for the human transcription factor E2F-2. E2F-2 is
CC involved in cell cycle regulation. In particular binding of E2F to
CC the retinoblastoma gene product (pRb) causes down regulation of the
CC transcription of any genes containing the E2F binding site. E2F-2 is
CC useful in the study of cell cycle regulation especially in the study
CC of pRb and certain viral oncogenes and oncoproteins.
XX
XX Sequence 2647 BP; 508 A; 861 C; 795 G; 483 T; 0 other;
Query Match 54.6%; Score 20.2; DB 16; Length 2647;
Best Local Similarity 75.8%; Pred. No. 1.1e+02;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
OY 5 atcaccatatacccttcccttgcgtggtg 37
Db 767 ATCCACTCTGATGCACCTCCCTTGGGGTGG 735
Search completed: June 4, 2002, 17:56:36
Job time: 7114 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 17:48:43 ; Search time 2137.99 Seconds
(without alignments)
362.154 Million cell updates/sec

Title: US-09-721-543A-16
Perfect score: 37
Sequence: 1 ctatatccaccatccccctccctgcgtggtggg 37

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- GenEmbl.*
1: gb_ba.*
2: gb_hgt.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pi.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pi.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	DB	ID	Description
NO.							

1	24.4	65.9	180313	2	AC098141	AC098141 Rattus no	
2	23.8	64.3	40205	9	AF045450	AF045450 Homo sapi	
3	64.3	142742	9	AF121782	AF121782 Homo sapi		
4	23.8	64.3	165382	9	AF064860	AF064860 Homo sapi	
5	23.8	64.3	203715	2	AC018807	AC018807 Homo sapi	
6	23.8	64.3	340000	9	HS21C080	AL163280 Homo sapi	
7	23.2	62.7	140000	9	AP001791	AP001791 Homo sapi	
c	8	23.2	62.7	141605	2	AC013732	AC013732 Homo sapi
c	9	23.2	62.7	148332	9	AC079586	AC079586 Homo sapi
c	10	23.2	62.7	162301	2	AC011023	AC011023 Homo sapi
11	23.2	62.7	180446	2	AP001597	AP001597 Homo sapi	
12	22.6	61.1	128481	2	AC094327	AC094327 Rattus no	
c	13	22.6	61.1	163338	9	AL162426	Human DNA
c	14	22.6	61.1	172588	2	AC007430	Homo sapi
c	15	22.6	61.1	236135	2	AC093559	Rattus no
c	16	22.4	60.5	101559	2	AC095646	Rattus no
17	22.4	60.5	170359	9	AC018545	Homo sapi	
c	18	22.4	60.5	198653	2	AC091474	Homo sapi
19	22.2	60.0	37959	2	AC102828	Mus muscu	
20	22.2	60.0	49804	2	AC101896	Mus muscu	
21	22.2	60.0	69900	2	AC101433	Mus muscu	
22	22.2	60.0	75693	2	AC026569	Mus muscu	
23	22.2	60.0	154062	9	AC022394	Homo sapi	
c	24	22.2	60.0	164946	9	AP000553	Homo sapi
25	22.2	60.0	166447	9	AC018751	Homo sapi	
c	26	22.2	60.0	169237	9	AC009516	Homo sapi
27	22.2	60.0	172989	9	AC096736	Homo sapi	
28	22.2	60.0	180144	2	AC094163	Rattus no	
29	21.8	58.9	1210	10	RATCRTXN	L15011 Rattus norv	
c	30	21.8	58.9	39984	9	AP000546	Homo sapi
c	31	21.8	58.9	69251	2	AC105064	Mus muscu
c	32	21.8	58.9	86519	9	AP001860	Homo sapi
c	33	21.8	58.9	110000	9	HSY237C10_1	Continuation (2 of
c	34	21.8	58.9	126480	2	AC040954	Mus muscu
c	35	21.8	58.9	129240	9	AC083826	Homo sapi
c	36	21.8	58.9	147521	2	AC013680	Homo sapi
c	37	21.8	58.9	157606	2	AC069172	Homo sapi
38	21.8	58.9	161973	9	HUAC002038	Homo sapi	
39	21.8	58.9	175878	2	AC097080	Rattus no	
40	21.8	58.9	187681	2	AC080124	Homo sapi	
41	21.8	58.9	188872	9	AC020750	Homo sapi	
c	42	21.8	58.9	206192	9	AL133173	Human DNA
c	43	21.6	58.4	4504	10	MMRPL	Y07941 M.musculus
c	44	21.6	58.4	122416	9	AL451050	Human DNA
c	45	21.6	58.4	146187	2	AC025854	Homo sapi

ALIGNMENTS

RESULT	1
AC098141	
LOCUS	180313 bp DNA linear HTG 21-DEC-2001
DEFINITION	Rattus norvegicus clone CH230-175N2, *** SEQUENCING IN PROGRESS
ACCESSION	AC098141
VERSION	AC098141.2 GI:17975707
KEYWORDS	HTG; HTGS_PHASE1.
SOURCE	Norway rat.
ORGANISM	Rattus norvegicus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
	Rattus.
REFERENCE	1 (bases 1 to 180313)
AUTHORS	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
	Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J.,
	Benton,J., Blinage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
	Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
	Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
	Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
	Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
	Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
	Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,

Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,
Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Hollaway, C.,
Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
Lewis, L., Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W.,
Louisege, H., Lozado, R., Lu, X., Lucier, A., Lucier, R., Luna, R.,
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
Martinez, E., Massey, M., Mawhiney, E., McLeod, M.P., Meador, M.,
Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S.,
Ogulu, M., Okwundu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M.,
Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N.,
Slisdon, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H.,
Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,
Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S.,
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 180313)
Worley, K.C.
Direct Submission
Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 21, 2001 this sequence version replaced gi:16327847.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GGMO
Center clone name: CH230-175N2
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 152722 bases at least Q40
Consensus quality: 158678 bases at least Q30
Consensus quality: 163556 bases at least Q20
Estimated insert size: 155476; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 2.6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 64 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved
* 1 11538: contig of 11538 bp in length
* 11539 11638: gap of unknown length
* 24157 24157: contig of 12519 bp in length
* 24158 24257: gap of unknown length
* 24258 29851: contig of 5594 bp in length
* 29852 29951: gap of unknown length
* 29952 35087: contig of 5136 bp in length
* 35088 35187: gap of unknown length
* 35188 40280: contig of 5093 bp in length
* 40281 40380: gap of unknown length
* 40381 46334: contig of 5954 bp in length
* 46335 51283: contig of 4849 bp in length
* 51284 55429: contig of 4046 bp in length
* 55430 59555: contig of 4026 bp in length
* 59556 64002: contig of 4347 bp in length
* 64003 67138: contig of 3036 bp in length
* 67139 71909: contig of 4670 bp in length
* 71910 76013: contig of 4004 bp in length
* 76014 79851: contig of 3739 bp in length
* 79852 82810: contig of 2859 bp in length
* 82811 85784: contig of 2873 bp in length
* 85785 89323: contig of 3440 bp in length
* 89324 92464: contig of 3041 bp in length
* 92465 96067: contig of 3503 bp in length
* 96068 98010: contig of 1843 bp in length
* 98011 101382: contig of 3272 bp in length
* 101383 104070: contig of 2588 bp in length
* 104071 106714: contig of 2544 bp in length
* 106715 106815: contig of 2845 bp in length
* 106816 109659: contig of 2185 bp in length
* 109660 111944: contig of 2858 bp in length
* 111945 112044: contig of 2858 bp in length
* 112045 115002: contig of 2033 bp in length
* 115003 117035: contig of 1577 bp in length
* 117036 118712: contig of 1099 bp in length
* 118713 119911: contig of 2483 bp in length
* 119912 122011: contig of 2480 bp in length
* 122012 122591: contig of 1783 bp in length
* 122592 124374: contig of 2150 bp in length
* 124375 126624: contig of 2483 bp in length
* 126625 129207: contig of 2257 bp in length
* 129208 131564: contig of 1944 bp in length
* 131565 133608: contig of 1584 bp in length
* 133609 135292: contig of 1445 bp in length
* 135293 136837: contig of 2144 bp in length
* 136838 139081: contig of 1132 bp in length
* 139082 140313: contig of 2393 bp in length
* 140314 142806: contig of 2393 bp in length
* 142807 144041: contig of 2393 bp in length
* 144042 145938: contig of 2393 bp in length
* 145939 147835: contig of 2393 bp in length
* 147836 149722: contig of 2393 bp in length
* 149723 151609: contig of 2393 bp in length
* 151610 153496: contig of 2393 bp in length
* 153497 155383: contig of 2393 bp in length
* 155384 157269: contig of 2393 bp in length
* 157270 159156: contig of 2393 bp in length
* 159157 161043: contig of 2393 bp in length
* 161044 162929: contig of 2393 bp in length
* 162930 164816: contig of 2393 bp in length
* 164817 166703: contig of 2393 bp in length
* 166704 168589: contig of 2393 bp in length
* 168590 170476: contig of 2393 bp in length
* 170477 172363: contig of 2393 bp in length
* 172364 174249: contig of 2393 bp in length
* 174250 176136: contig of 2393 bp in length
* 176137 178022: contig of 2393 bp in length
* 178023 179909: contig of 2393 bp in length
* 179910 181795: contig of 2393 bp in length
* 181796 183682: contig of 2393 bp in length
* 183683 185569: contig of 2393 bp in length
* 185570 187455: contig of 2393 bp in length
* 187456 189342: contig of 2393 bp in length
* 189343 191229: contig of 2393 bp in length
* 191230 193115: contig of 2393 bp in length
* 193116 195002: contig of 2393 bp in length
* 195003 196889: contig of 2393 bp in length
* 196890 198775: contig of 2393 bp in length
* 198776 200662: contig of 2393 bp in length
* 200663 202549: contig of 2393 bp in length
* 202550 204435: contig of 2393 bp in length
* 204436 206322: contig of 2393 bp in length
* 206323 208208: contig of 2393 bp in length
* 208209 210095: contig of 2393 bp in length
* 210096 211981: contig of 2393 bp in length
* 211982 213868: contig of 2393 bp in length
* 213869 215754: contig of 2393 bp in length
* 215755 217641: contig of 2393 bp in length
* 217642 219527: contig of 2393 bp in length
* 219528 221413: contig of 2393 bp in length
* 221414 223300: contig of 2393 bp in length
* 223301 225186: contig of 2393 bp in length
* 225187 227072: contig of 2393 bp in length
* 227073 228958: contig of 2393 bp in length
* 228959 230845: contig of 2393 bp in length
* 230846 232731: contig of 2393 bp in length
* 232732 234618: contig of 2393 bp in length
* 234619 236504: contig of 2393 bp in length
* 236505 238391: contig of 2393 bp in length
* 238392 240277: contig of 2393 bp in length
* 240278 242163: contig of 2393 bp in length
* 242164 244049: contig of 2393 bp in length
* 244050 245936: contig of 2393 bp in length
* 245937 247822: contig of 2393 bp in length
* 247823 249708: contig of 2393 bp in length
* 249709 251595: contig of 2393 bp in length
* 251596 253481: contig of 2393 bp in length
* 253482 255368: contig of 2393 bp in length
* 255369 257253: contig of 2393 bp in length
* 257254 259140: contig of 2393 bp in length
* 259141 261026: contig of 2393 bp in length
* 261027 262913: contig of 2393 bp in length
* 262914 264800: contig of 2393 bp in length
* 264801 266686: contig of 2393 bp in length
* 266687 268573: contig of 2393 bp in length
* 268574 270459: contig of 2393 bp in length
* 270460 272346: contig of 2393 bp in length
* 272347 274232: contig of 2393 bp in length
* 274233 276119: contig of 2393 bp in length
* 276120 278005: contig of 2393 bp in length
* 278006 279892: contig of 2393 bp in length
* 279893 281778: contig of 2393 bp in length
* 281779 283664: contig of 2393 bp in length
* 283665 285551: contig of 2393 bp in length
* 285552 287437: contig of 2393 bp in length
* 287438 289323: contig of 2393 bp in length
* 289324 291209: contig of 2393 bp in length
* 291210 293096: contig of 2393 bp in length
* 293097 294982: contig of 2393 bp in length
* 294983 296868: contig of 2393 bp in length
* 296869 298754: contig of 2393 bp in length
* 298755 300641: contig of 2393 bp in length
* 300642 302527: contig of 2393 bp in length
* 302528 304413: contig of 2393 bp in length
* 304414 306299: contig of 2393 bp in length
* 306300 308186: contig of 2393 bp in length
* 308187 310072: contig of 2393 bp in length
* 310073 311958: contig of 2393 bp in length
* 311959 313844: contig of 2393 bp in length
* 313845 315731: contig of 2393 bp in length
* 315732 317617: contig of 2393 bp in length
* 317618 319504: contig of 2393 bp in length
* 319505 321389: contig of 2393 bp in length
* 321390 323275: contig of 2393 bp in length
* 323276 325162: contig of 2393 bp in length
* 325163 327048: contig of 2393 bp in length
* 327049 328933: contig of 2393 bp in length
* 328934 330820: contig of 2393 bp in length
* 330821 332706: contig of 2393 bp in length
* 332707 334592: contig of 2393 bp in length
* 334593 336478: contig of 2393 bp in length
* 336479 338364: contig of 2393 bp in length
* 338365 340250: contig of 2393 bp in length
* 340251 342136: contig of 2393 bp in length
* 342137 344022: contig of 2393 bp in length
* 344023 345908: contig of 2393 bp in length
* 345909 347794: contig of 2393 bp in length
* 347795 349680: contig of 2393 bp in length
* 349681 351566: contig of 2393 bp in length
* 351567 353452: contig of 2393 bp in length
* 353453 355338: contig of 2393 bp in length
* 355339 357224: contig of 2393 bp in length
* 357225 359110: contig of 2393 bp in length
* 359111 361000: contig of 2393 bp in length
* 361001 362886: contig of 2393 bp in length
* 362887 364772: contig of 2393 bp in length
* 364773 366658: contig of 2393 bp in length
* 366659 368544: contig of 2393 bp in length
* 368545 370430: contig of 2393 bp in length
* 370431 372316: contig of 2393 bp in length
* 372317 374202: contig of 2393 bp in length
* 374203 376088: contig of 2393 bp in length
* 376089 377973: contig of 2393 bp in length
* 377974 379859: contig of 2393 bp in length
* 379860 381746: contig of 2393 bp in length
* 381747 383631: contig of 2393 bp in length
* 383632 385517: contig of 2393 bp in length
* 385518 387403: contig of 2393 bp in length
* 387404 389289: contig of 2393 bp in length
* 389290 391175: contig of 2393 bp in length
* 391176 393061: contig of 2393 bp in length
* 393062 394947: contig of 2393 bp in length
* 394948 396833: contig of 2393 bp in length
* 396834 398718: contig of 2393 bp in length
* 398719 400604: contig of 2393 bp in length
* 400605 402489: contig of 2393 bp in length
* 402490 404375: contig of 2393 bp in length
* 404376 406260: contig of 2393 bp in length
* 406261 408146: contig of 2393 bp in length
* 408147 410032: contig of 2393 bp in length
* 410033 411917: contig of 2393 bp in length
* 411918 413803: contig of 2393 bp in length
* 413804 415688: contig of 2393 bp in length
* 415689 417574: contig of 2393 bp in length
* 417575 419459: contig of 2393 bp in length
* 419460 421345: contig of 2393 bp in length
* 421346 423230: contig of 2393 bp in length
* 423231 425115: contig of 2393 bp in length
* 425116 427001: contig of 2393 bp in length
* 427002 428886: contig of 2393 bp in length
* 428887 430771: contig of 2393 bp in length
* 430772 432657: contig of 2393 bp in length
* 432658 434542: contig of 2393 bp in length
* 434543 436428: contig of 2393 bp in length
* 436429 438313: contig of 2393 bp in length
* 438314 440198: contig of 2393 bp in length
* 440199 442084: contig of 2393 bp in length
* 442085 443969: contig of 2393 bp in length
* 443970 445855: contig of 2393 bp in length
* 445856 447740: contig of 2393 bp in length
* 447741 449626: contig of 2393 bp in length
* 449627 451511: contig of 2393 bp in length
* 451512 453396: contig of 2393 bp in length
* 453397 455282: contig of 2393 bp in length
* 455283 457167: contig of 2393 bp in length
* 457168 459052: contig of 2393 bp in length
* 459053 460937: contig of 2393 bp in length
* 460938 462822: contig of 2393 bp in length
* 462823 464708: contig of 2393 bp in length
* 464709 466592: contig of 2393 bp in length
* 466593 468477: contig of 2393 bp in length
* 468478 470363: contig of 2393 bp in length
* 470364 472248: contig of 2393 bp in length
* 472249 474133: contig of 2393 bp in length
* 474134 476018: contig of 2393 bp in length
* 476019 477903: contig of 2393 bp in length
* 477904 479788: contig of 2393 bp in length
* 479789 481673: contig of 2393 bp in length
* 481674 483557: contig of 2393 bp in length
* 483558 485442: contig of 2393 bp in length
* 485443 487326: contig of 2393 bp in length
* 487327 489211: contig of 2393 bp in length
* 489212 491095: contig of 2393 bp in length
* 491096 492979: contig of 2393 bp in length
* 492980 494864: contig of 2393 bp in length
* 494865 496748: contig of 2393 bp in length
* 496749 498633: contig of 2393 bp in length
* 498634 500517: contig of 2393 bp in length
* 500518 502402: contig of 2393 bp in length
* 502403 504286: contig of 2393 bp in length
* 504287 506170: contig of 2393 bp in length
* 506171 508055: contig of 2393 bp in length
* 508056 509939: contig of 2393 bp in length
* 509940 511824: contig of 2393 bp in length
* 511825 513708: contig of 2393 bp in length
* 513709 515592: contig of 2393 bp in length
* 515593 517476: contig of 2393 bp in length
* 517477 519360: contig of 2393 bp in length
* 519361 521244: contig of 2393 bp in length
* 521245 523128: contig of 2393 bp in length
* 523129 525012: contig of 2393 bp in length
* 525013 526896: contig of 2393 bp in length
* 526897 528779: contig of 2393 bp in length
* 528780 530663: contig of 2393 bp in length
* 530664 532547: contig of 2393 bp in length
* 532548 534431: contig of 2393 bp in length
* 534432 536315: contig of 2393 bp in length
* 536316 538198: contig of 2393 bp in length
* 538199 540082: contig of 2393 bp in length
* 540083 541965: contig of 2393 bp in length
* 541966 543848: contig of 2393 bp in length
* 543849 545732: contig of 2393 bp in length
* 545733 547615: contig of 2393 bp in length
* 547616 549498: contig of 2393 bp in length
* 549499 551381: contig of 2393 bp in length
* 551382 553264: contig of 2393 bp in length
* 553265 555147: contig of 2393 bp in length
* 555148 557030: contig of 2393 bp in length
* 557031 558913: contig of 2393 bp in length
* 558914 560796: contig of 2393 bp in length
* 560797 562679: contig of 2393 bp in length
* 562680 564562: contig of 2393 bp in length
* 564563 566445: contig of 2393 bp in length
* 566446 568328: contig of 2393 bp in length
* 568329 570211: contig of 2393 bp in length
* 570212 572094: contig of 2393 bp in length
* 572095 573977: contig of 2393 bp in length
* 573978 575860: contig of 2393 bp in length
* 575861 577743: contig of 2393 bp in length
* 577744 579626: contig of 2393 bp in length
* 579627 581509: contig of 2393 bp in length
* 581510 583391: contig of 2393 bp in length
* 583392 585274: contig of 2393 bp in length
* 585275 587157: contig of 2393 bp in length
* 587158 589039: contig of 2393 bp in length
* 589040 590922: contig of 2393 bp in length
* 590923 592804: contig of 2393 bp in length
* 592805 594687: contig of 2393 bp in length
* 594688 596569: contig of 2393 bp in length
* 596570 598452: contig of 2393 bp in length
* 598453 600334: contig of 2393 bp in length
* 600335 602217: contig of 2393 bp in length
* 602218 604100: contig of 2393 bp in length
* 604101 605982: contig of 2393 bp in length
* 605983 607865: contig of 2393 bp in length
* 607866 609748: contig of 2393 bp in length
* 609749 611630: contig of 2393 bp in length
* 611631 613513: contig of 2393 bp in length
* 613514 615395: contig of 2393 bp in length
* 615396 617277: contig of 2393 bp in length
* 617278 619159: contig of 2393 bp in length
* 619160 621042: contig of 2393 bp in length
* 621043 622923: contig of 2393 bp in length
* 622924 624805: contig of 2393 bp in length
* 624806 626687: contig of 2393 bp in length
* 626688 628569: contig of 2393 bp in length
* 628570 630451: contig of 2393 bp in length
* 630452 632333: contig of 2393 bp in length
* 632334 634214: contig of 2393 bp in length
* 634215 636096: contig of 2393 bp in length
* 636097 637977: contig of 2393 bp in length
* 637978 639858: contig of 2393 bp in length
* 639859 641739: contig of 2393 bp in length
* 641740 643621: contig of 2393 bp in length
* 643622 645502: contig of 2393 bp in length
* 645503 647383: contig of 2393 bp in length
* 647384 649264: contig of 2393 bp in length
* 649265 651145: contig of 2393 bp in length
* 651146 653026: contig of 2393 bp in length
* 653027 654907: contig of 2393 bp in length
* 654908 656788: contig of 2393 bp in length
* 656789 658669: contig of 2393 bp in length
* 658670 660550: contig of 2393 bp in length
* 660551 662431: contig of 2393 bp in length
* 662432 664312: contig of 2393 bp in length
* 664313 666193: contig of 2393 bp in length
* 666194 668074: contig of 2393 bp in length
* 668075 669955: contig of 2393 bp in length
* 669956 671836: contig of 2393 bp in length
* 671837 673717: contig of 2393 bp in length
* 673718 675598: contig of 2393 bp in length
* 675599 677479: contig of 2393 bp in length
* 677480 679360: contig of 2393 bp in length
* 679361 681241: contig of 2393 bp in length
* 681242 683122: contig of 2393 bp in length
* 683123 685003: contig of 2393 bp in length
* 685004 686884: contig of 2393 bp in length
* 686885 688765: contig of 2393 bp in length
* 688766 690646: contig of 2393 bp in length
* 690647 692527: contig of 2393 bp in length
* 692528 694408: contig of 2393 bp in length
* 694409 696289: contig of 2393 bp in length
* 696290 698170: contig of 2393 bp in length
* 698171 700051: contig of 2393 bp in length
* 700052 701932: contig of 2393 bp in length
* 701933 703813: contig of 2393 bp in length
* 703814 705694: contig of 2393 bp in length
* 705695 707575: contig of 2393 bp in length
* 707576 709456: contig of 2393 bp in length
* 709457 711337: contig of 2393 bp in length
* 711338 713218: contig of 2393 bp in length
* 713219 715099: contig of 2393 bp in length
* 715100 716979: contig of 2393 bp in length
* 716980 718860: contig of 2393 bp in length
* 718861 720740: contig of 2393 bp in length
* 720741 722621: contig of 2393 bp in length
* 722622 724501: contig of 2393 bp in length
* 724502 726382: contig of 2393 bp in length
* 726383 728262: contig of 2393 bp in length
*


```

repeat_region 11478..11597
/evidence-not_experimental
/rpt_family="L2"
exon 11570..11704
/note="GRAIL, score = 99.000%, comment = excellent shadow"
/evidence-not_experimental
exon 11570..11704
/note="M2EF, score = 66.1%"
/evidence-not_experimental
complement(11687..11735)
/rpt_family="MIR"
/evidence-not_experimental
complement(11774..12320)
/rpt_family="MLTIE"
/evidence-not_experimental
complement(11828..11949)
/note="GRAIL, score = 50.000%, comment = good"
/evidence-not_experimental
12823..12843
/rpt_family="AluSx"
/evidence-not_experimental
13102..13457
/rpt_family="THE1B"
/evidence-not_experimental
complement(13481..13519)
/note="GRAIL, score = 65.000%, comment = good"
/evidence-not_experimental
15065..15188
/rpt_family="HERVL"
/evidence-not_experimental
15130..15378
/note="GC score = 10.60 (249bp)"
/note="Region: GC content"
/evidence-not_experimental
15369..16268
/rpt_family="MERVL"
/evidence-not_experimental
complement(16623..16714)
/note="GRAIL, score = 72.000%, comment = good"
/evidence-not_experimental
16661..16730
/note="M2EF, score = 93.5%"
/evidence-not_experimental
16880..17318
/rpt_family="MLTIC"
/evidence-not_experimental
17326..17436
/rpt_family="LTRL6C"
/evidence-not_experimental
complement(17540..17675)
/rpt_family="MIR"
/evidence-not_experimental
complement(17805..17864)
/rpt_family="MLTIC"
/evidence-not_experimental
18076..18098
/note="GRAIL, score = 45.000%, comment = marginal shadow"
/evidence-not_experimental
complement(18219..18349)
/note="GenScan, score = 2.99%, comment = Internal_exon 131
bp frame: 1 phase: 2"
/evidence-not_experimental
complement(18717..19211)
/rpt_family="MLTIC"
/evidence-not_experimental
complement(19252..19399)
/note="GRAIL, score = 71.000%, comment = good"
/evidence-not_experimental
19269..19350
/note="GRAIL, score = 81.000%, comment = excellent shadow"
/evidence-not_experimental
complement(19379..19399)
/note="Xpound exon prediction, score = 60% (0%)"

```

```

exon /evidence-not_experimental
complement(19501..19881)
/note="M2EF, score = 50.6%"
/evidence-not_experimental
19455..20020
/note="homology = 100.00%, score = 38, counts = 2"
/rpt_type=tandem
/rpt_unit=ttcttcgacactctctgttgagaggtcggtttatcgc
/evidence-not_experimental
20223..20643
/rpt_family="MLTIE"
/evidence-not_experimental
20645..20910
/rpt_family="MLTIF"
/evidence-not_experimental
20961..21007
/rpt_family="MLTIF"
/evidence-not_experimental
21483..21781
/rpt_family="AluJb"
/evidence-not_experimental
complement(21504..21756)
/note="M2EF, score = 79.4%"
/evidence-not_experimental
complement(22325..22806)
/rpt_family="MLTIC"

Query Match 64.3%; Score 23.8; DB 9; Length 40205;
Best Local Similarity 80.0%; Pred. No. 60;
Matches 28; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 3 atatccaccatcccccttccttcgtgtggg 37
||||| ||||| ||||| ||||| |||||
Db 20137 ATGTCCTCCATATTCCTTCTCTTGAGTGGG 20171

RESULT 3
AF121782
LOCUS Homo sapiens chromosome 21q22.3 PAC 206A10, complete sequence.
DEFINITION AF121782
ACCESSION AF121782
VERSION AF121782.1 GI:4210991
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 142742)
AUTHORS Taudien,S., Dagand,E., Hildmann,T., Nordstiek,G., Drescher,B.,
Schattevoy,R., Weber,J., Schilling,M., Menzel,U., Yaspo,M.L. and
Rosenthal,A.
TITLE Direct Submission
JOURNAL Submitted (21-JAN-1999) Genome Analysis, Institute for Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES
source
1..142742
/organism="Homo sapiens"
/db_xref="taxon.9606"
/chromosome="21"
/map="21q22.3"
/clone="PAC 206A10"
3..118
/rpt_family="MIR"
/evidence-not_experimental
complement(140..415)
/rpt_family="MLT2E"
/evidence-not_experimental
complement(668..881)
/rpt_family="MER58A"
/evidence-not_experimental
complement(1078..1162)
/note="GRAIL, score = 86.000%, comment = excellent shadow"
/evidence-not_experimental

```



```

/exvidence-not_experimental
18352..18420
/notes="GRAIL, score = 51.000%, comment = good shadow"
/exvidence-not_experimental
complement(18397..18584)
/notes="GRAIL, score = 66.000%, comment = good"
/exvidence-not_experimental
complement(18643..24647)
/rpt_family="HERV17"
/exvidence-not_experimental
complement(19228..19361)
/notes="M2EF, score = 92.8%"
/exvidence-not_experimental
20900..21135
/notes="M2EF, score = 51.4%"
/exvidence-not_experimental
21150..21275
/notes="Xpound exon prediction, score = 74% (0%)"
/exvidence-not_experimental

Query Match      64.3%; Score 23.8; DB 9; Length 142742;
Best Local Similarity 80.0%; Pred. No. 50;
Matches 28; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 atataccaccatccccctcccttcgctggtgg 37
Db 10158 ATGCTCTTCATATTCCTCTCTTTGAGTGTGG 10192

RESULT 4
AF064860 165382 bp DNA linear PRI 02-JUN-1998
LOCUS Homo sapiens chromosome 21q22.3 PAC 70124, complete sequence.
DEFINITION AF064860
ACCESSION AF064860
VERSION AF064860.1 GI:3171153
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 165382)
AUTHORS Rump,A., Dagand,E., Hildmann,T., Nordsiek G., Drescher,B.,
Weber,J., Schattevoy,R., Yaspo,M.-L. and Rosenthal,A.
TITLE Direct Submission
JOURNAL Submitted (12-MAY-1998) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES
source
1..165382
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="21q22.3"
complement(<1..77)
/notes="GenScan, score = 3.23%, comment = Single-exon_gene
279 bp frame: 2 phase: 0"
/exvidence-not_experimental
complement(43..366)
/rpt_family="Ltr16A"
/exvidence-not_experimental
complement(485..648)
/rpt_family="MER63B"
/exvidence-not_experimental
complement(977..1626)
/rpt_family="HERV1"
/exvidence-not_experimental
complement(1708..2125)
/rpt_family="HERV1"
/exvidence-not_experimental
complement(2836..3210)
/rpt_family="MLT1A1"
/exvidence-not_experimental
2928..3064
/notes="GRAIL, score = 64.000%, comment = good"

```

```

/exvidence-not_experimental
complement(3456..3559)
/rpt_family="L2"
/exvidence-not_experimental
4328..4468
/rpt_family="MER86"
/exvidence-not_experimental
5138..5597
/rpt_family="L1M4"
/exvidence-not_experimental
complement(5641..5850)
/rpt_family="MIR"
/exvidence-not_experimental
complement(6246..6452)
/rpt_family="Ltr10A"
/exvidence-not_experimental
complement(6763..6923)
/rpt_family="MIR"
/exvidence-not_experimental
7437..7538
/rpt_family="MSTA"
/exvidence-not_experimental
7542..8239
/rpt_family="Ltr8"
/exvidence-not_experimental
8242..8546
/rpt_family="MSTA"
/exvidence-not_experimental
complement(8677..8727)
/rpt_family="L2"
/exvidence-not_experimental
9938..10445
/rpt_family="MLT1D"
/exvidence-not_experimental
10148..10273
/notes="GRAIL, score = 50.000%, comment = good"
/exvidence-not_experimental
10451..10538
/rpt_family="MIR"
/exvidence-not_experimental
complement(10682..11042)
/rpt_family="MER7A"
/exvidence-not_experimental
11815..11944
/notes="homology = 99.20%, score = 123, counts = 26"
/rpt_type=tandem
/rpt_unit=tcctc
/exvidence-not_experimental
complement(11944..12237)
/rpt_family="AluSc"
/exvidence-not_experimental
complement(11973..12013)
/notes="Xpound exon prediction, score = 83% (0%)"
/exvidence-not_experimental
complement(12139..12170)
/notes="Xpound exon prediction, score = 77% (0%)"
/exvidence-not_experimental
12770..14551
/rpt_family="L1M45"
/exvidence-not_experimental
13473..13620
/notes="GRAIL, score = 51.000%, comment = good"
/exvidence-not_experimental
14602..14648
/notes="M2EF, score = 91.1%"
/exvidence-not_experimental
complement(14681..14973)
/rpt_family="AluSg"
/exvidence-not_experimental
complement(15289..15406)
/notes="GRAIL, score = 58.000%, comment = good"
/exvidence-not_experimental
complement(15297..15346)

```

```
/note="Xpound exon prediction, score = 87% (0%)"
/evidence-not_experimental
15436..16003
/rpt_family="LIME"
/evidence-not_experimental
complement(16008..16227)
/rpt_family="MLT1A1"
/evidence-not_experimental
complement(16248..16421)
/rpt_family="LIPB3"
/evidence-not_experimental
16259..16359
/note="MZEf, score = 72.4%"
/evidence-not_experimental
complement(16283..16406)
/note="MZEf, score = 61%"
/evidence-not_experimental
complement(16429..16576)
/rpt_family="MLT1A1"
/evidence-not_experimental
16577..17146
/rpt_family="LIM4"
/evidence-not_experimental
17153..17451
/rpt_family="AluSc"
/evidence-not_experimental
17453..18912
/rpt_family="LIM4"
/evidence-not_experimental
17486..17522
/note="MZEf, score = 70.3%"
/evidence-not_experimental
18775..18868
/note="MZEf, score = 62.5%"
/evidence-not_experimental
18921..18984
/rpt_family="LIMC/D"
/evidence-not_experimental
18990..19261
/rpt_family="AluJb"
/evidence-not_experimental
19262..19415
/rpt_family="LIMC/D"
/evidence-not_experimental
complement(19413..19788)
/rpt_family="LIPAL3"
/evidence-not_experimental
19820..20234
/rpt_family="LIMC/D"
/evidence-not_experimental
20293..20572
/rpt_family="LIMC/D"
/evidence-not_experimental
complement(20623..20924)
/rpt_family="AluJb"
/evidence-not_experimental
20810..20978
/note="MZEf, score = 62.6%"
/evidence-not_experimental
complement(21141..21424)
/rpt_family="AluJo"
/evidence-not_experimental
22062..22191
/rpt_family="LIPB1"
/evidence-not_experimental
22432..22504
/note="GAIL, score = 59.000%, comment = good"
/evidence-not_experimental
22572..22795
/rpt_family="MLT2"
/evidence-not_experimental
complement(22810..23112)
/rpt_family="AluJo"
```

```
/evidence-not_experimental
23130..23282
/rpt_family="MLT2"
/evidence-not_experimental
23690..23750
/note="MZEf, score = 95.1%"
/evidence-not_experimental
23906..24295
/rpt_family="MLT1H"
/evidence-not_experimental
24048..24211
/note="GAIL, score = 52.000%, comment = good"

Query Match      64.3%; Score 23.8; DB 9; Length 165382;
Best Local Similarity 80.0%; Pred. No. 48;
Matches 28; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 3 atatcaccatcatcccttcccttgcgtggtgg 37
DB 156149 ATGCTCTCCATATCCCTTCTCTTTGACTGTGGG 156183

RESULT 5
AC018807
LOCUS      203715 bp DNA linear HTG 07-JUL-2000
DEFINITION Homo sapiens chromosome 8 clone RP11-648J12, WORKING DRAFT
AC018807
VERSION    AC018807.5 GI:8748944
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 203715)
            Waterston,R.H.
            The sequence of Homo sapiens clone
            Unpublished
REFERENCE  2 (bases 1 to 203715)
            Waterston,R.H.
            Direct Submission
            Submitted (03-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT    On Jun 27, 2000 this sequence version replaced gi:7235308.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0648J12
----- Summary Statistics -----
Sequencing vector: M13; 69%
Chemistry: Dye-terminator Big Dye; 31% of reads
Chemistry: Dye-terminator Big Dye; 31% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 191142 bases at least Q40
Consensus quality: 195733 bases at least Q30
Consensus quality: 197952 bases at least Q20
Insert size: 190000; agarose-gel
Quality coverage: 4.28 in Q20 bases; agarose-gel
Quality coverage: 4.08 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
```

* be preserved.

```

1
1192: contig of 1192 bp in length
1292: gap of unknown length
2406: contig of 1114 bp in length
2506: gap of unknown length
3757: contig of 1251 bp in length
3857: gap of unknown length
4952: contig of 1095 bp in length
5052: gap of unknown length
6760: contig of 1708 bp in length
6860: gap of unknown length
8708: contig of 1846 bp in length
8808: gap of unknown length
9939: contig of 1133 bp in length
10039: gap of unknown length
11704: contig of 1665 bp in length
11804: gap of unknown length
13577: contig of 1773 bp in length
13677: gap of unknown length
15044: contig of 1367 bp in length
15144: gap of unknown length
15145: contig of 1526 bp in length
16670: gap of unknown length
16771: contig of 1761 bp in length
18531: contig of 1761 bp in length
18631: gap of unknown length
19872: contig of 1241 bp in length
19972: gap of unknown length
24360: contig of 4388 bp in length
24460: gap of unknown length
28500: contig of 4040 bp in length
28600: gap of unknown length
33938: contig of 5338 bp in length
34038: gap of unknown length
40567: contig of 6529 bp in length
40667: gap of unknown length
46244: contig of 5577 bp in length
46344: gap of unknown length
54700: contig of 8356 bp in length
54800: gap of unknown length
62611: contig of 7811 bp in length
62711: gap of unknown length
68823: contig of 24112 bp in length
86923: gap of unknown length
109124: contig of 22219 bp in length
109242: gap of unknown length
138612: contig of 29370 bp in length
138613: gap of unknown length
138713: contig of 29959 bp in length
168713: contig of 29959 bp in length
168772: gap of unknown length
168772: contig of 34944 bp in length.

```

FEATURES

source

```

Location/Qualifiers
1. .203715
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/clone="RP11-648J12"
1. .1192
/notes="assembly_name:Contig24"
1293. .2406
/notes="assembly_name:Contig29"
2507. .3757
/notes="assembly_name:Contig31"
3858. .4952
/notes="assembly_name:Contig35"
5053. .6760
/notes="assembly_name:Contig50"
6861. .8706
/notes="assembly_name:Contig52"
8807. .9939
/notes="assembly_name:Contig53"
10040. .11704
/notes="assembly_name:Contig55"
11805. .13577

Query Match          64.3%   Score 23.8;   DB 2;   Length 203715;
Best Local Similarity 80.0%;   Pred. No. 47;
Matches 28;   Conservative 0;   Mismatches 7;   Indels 0;   Gaps 0;

QY  2  tatatccaccatcccccttcccccttgcgtgtgg 36
      ||||| || ||||| ||||| || |||||
Db  97755 TATATCCTCATTATCCCTTCCCTAGCCCTCG 97789

RESULT 6
HS21C080
LOCUS      HS21C080          340000 bp   DNA    linear    PRI 24-MAY-2000
DEFINITION Homo sapiens chromosome 21 segment HS21C080.
ACCESSION  AL163280 AP001735 BA000005
VERSION     AL163280.2 GI:7717369
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 340000)
Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyma,S.,
Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G.,
Hornischer,K., Brandt,P., Schaefer,M., Schoen,O., Desario,A.,
Reichelt,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S.,
Hennig,S., Riesemann,L., Dagand,E., Wehrmeyer,S., Borzym,K.,
Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,K. and
Yaspo,M.L.

```

```

misc_feature      /note="assembly_name:Contig56"
13678. .15044
/notes="assembly_name:Contig57"
15145. .16670
/notes="assembly_name:Contig58"
16771. .18531
/notes="assembly_name:Contig59"
18632. .19872
/notes="assembly_name:Contig60"
19973. .24360
/notes="assembly_name:Contig61"
24461. .28500
clone_end:SP6
vector_side:right"
28601. .33938
/notes="assembly_name:Contig63"
34039. .40567
/notes="assembly_name:Contig64"
40668. .46244
/notes="assembly_name:Contig65"
46345. .54700
/notes="assembly_name:Contig66"
54801. .62611
/notes="assembly_name:Contig67"
62712. .68823
/notes="assembly_name:Contig68"
68924. .109142
/notes="assembly_name:Contig69"
109243. .138612
/notes="assembly_name:Contig70"
138713. .168671
/notes="assembly_name:Contig71"
168772. .203715
/notes="assembly_name:Contig72"
BASE COUNT  66684 a 34739 c 35439 g 64441 t 2412 others
ORIGIN

```


TITLE JOURNAL	Direct Submission Submitted (05-MAY-2000) The Chromosome 21 Mapping and Sequencing Consortium: * RIKEN Genomic Sciences Center, Human Genome Research Group * Institute of Molecular Biotechnology, Genome Analysis * Keio University School of Medicine, Dept. of Molecular Biology * GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular Genetics (addresses see below) The Chromosome 21 Mapping and Sequencing Consortium consists of * RIKEN Genomic Sciences Center, Human Genome Research Group, * Sagami-hara 228-8555, Japan, * e.mail: sakaki@gscl.riken.go.jp * URL: http://hgp.gsc.riken.go.jp/ and * Institute of Molecular Biotechnology, Genome Analysis, * Beutenbergstrasse 11, D-07745 Jena, Germany, * e.mail: gscj-submit@genome.imb-jena.de * URL: http://genome.imb-jena.de/ and * Keio University School of Medicine, Dept. of Molecular Biology, * Tokyo 160-8582, Japan, * e.mail: shimizu@dbm-med.keio.ac.jp * URL: http://adenine.dmb.med.keio.ac.jp/ and * GBF, Dept. of Genome Analysis, * Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail: info.genome@gbf.de * URL: http://genome.gbf.de/ and * Max-Planck Institute for Molecular Genetics, * Innestrasse 73, D-14195 Berlin, Germany, * e.mail: info-chr21@molgen.mpg.de * URL: http://chr21.rz-berlin.mpg.de/ Location/Qualifiers 1. 340000		
FEATURES source	/organism="Homo sapiens" /db_xref="taxon:9606" /map="21q22.3" <1..125946 /note="Accession No. AF121897" /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="21" /map="21q22.3" /clone_lib="RPC11, 3-5 PAC library" 1088..1566 /note="LIM4"	repeat_region	join(<3473..3565,12649..12678,15331..15460,18739..19157) /gene="SH3BGR"
source	/rpt_family="LINE/L1" /rpt_type=DISPERSED 2398..2495 /note="(TCTA)n" /rpt_family="Simple_repeat" /rpt_type=TANDEM complement(2496..2878) /note="THE1C" /rpt_family="LTR/MaLR" /rpt_type=DISPERSED 2879..2931 /note="(TCTA)n" /rpt_family="Simple_repeat" /rpt_type=TANDEM 2932..3004 /note="(TCCA)n" /rpt_family="Simple_repeat" 3005..3074 /note="(TCCA)n" /rpt_family="Simple_repeat" 3473..3565 /gene="SH3BGR" /number=4	repeat_region	join(<3473..3565,12649..12678,15331..15426) /partial /gene="SH3BGR" /note="Accession No. X93498" /codon_start=1 /product="21-Glutamic Acid Rich protein 21-GARP" /protein_id="CAB90445.1" /db_xref="GI:7717370" /translation="GSEKAEGETEAQKESGSDVGNLPEAQKNEEGETATEETEE IAMEGAEGAEETETAEGETPEDEDS" 5542..5829 /note="AluSx" /rpt_family="SINE/Alu" /rpt_type=DISPERSED complement(6887..7067) /note="LIMD3" /rpt_family="LINE/L1" /rpt_type=DISPERSED complement(8482..8762) /note="AluJo" /rpt_family="SINE/Alu" /rpt_type=DISPERSED complement(8865..8964) /note="L2" /rpt_family="LINE/L2" /rpt_type=DISPERSED 9131..9424 /note="AluSq" /rpt_family="SINE/Alu" /rpt_type=DISPERSED complement(9723..9882) /note="L2" /rpt_family="LINE/L2" /rpt_type=DISPERSED complement(9989..10077) /note="L2" /rpt_family="LINE/L2" /rpt_type=DISPERSED complement(10078..10429) /note="THE1B" /rpt_family="LTR/MaLR" /rpt_type=DISPERSED complement(10430..10651) /note="L2" /rpt_family="LINE/L2" /rpt_type=DISPERSED 10980..11000 /note="(TTTA)n" /rpt_family="Simple_repeat" /rpt_type=TANDEM complement(11002..11282) /note="AluSp" /rpt_family="SINE/Alu" /rpt_type=DISPERSED complement(11315..11556) /note="AluJo" /rpt_family="SINE/Alu" /rpt_type=DISPERSED 11702..11753 /note="MIR" /rpt_family="SINE/MIR" /rpt_type=DISPERSED 12649..12678 /gene="SH3BGR" /number=5 complement(13115..13425) /note="AluSx" /rpt_family="SINE/Alu" /rpt_type=DISPERSED
exon		exon	

```

repeat_region complement(13598..13892)
/nc="AluY"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region 14013..14131
/nc="CA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
repeat_region 14561..14598
/nc="(TA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
repeat_region 14758..15215
/nc="(TG)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
exon 15331..15460
/gene="SH3BGR"
/number=6
3'UTR 15427..15460
/gene="SH3BGR"
repeat_region 16255..16471
/nc="AluSg/x"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region complement(18013..18312)
/nc="AluDb"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
exon 18739..19157
/gene="SH3BGR"

```

Query Match 64.3%; Score 23.8; DB 9; Length 340000;
 Best Local Similarity 80.0%; Pred. NO. 43;
 Matches 28; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

QY 3 ataccaccatccctccctccctggtggtgg 37
||||| ||||| ||||| ||||| ||||| |||||
Db 234299 ATGCTTCATATCCCTCTCTTGTGAGTG 234333

```

```

RESULT 7
AP001791
LOCUS AP001791 140000 bp DNA linear PRI 08-DEC-2001
DEFINITION Homo sapiens genomic DNA, chromosome 11q, clone:RP11-427G12,
complete sequence.
ACCESSION AP001791
VERSION AP001791.4 GI:17425233
KEYWORDS HTG.
SOURCE Homo sapiens DNA, clone:RP11-427G12.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 140000)
AUTHORS Fujiyama,A., Yada,T., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Hattori,M., Ishii,K., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (11-APR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsuri-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsr.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT On Dec 7, 2001 this sequence version replaced gi:9927282.
FEATURES
Location/Qualifiers
source 1..140000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-427G12"

```

```

BASE COUNT 45897 a 26034 c 24406 g 43663 t
ORIGIN

```

```

Query Match 62.7%; Score 23.2; DB 9; Length 140000;
Best Local Similarity 77.8%; Pred. NO. 83;
Matches 28; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

```

QY 2 tatatccacccatccccctccctggtggtgg 37
||||| ||||| ||||| ||||| ||||| |||||
Db 86668 TATATTCACCAATATCCACTCTCTCTACTGTGGG 86703

```

```

RESULT 8
AC013732/c
LOCUS AC013732 141605 bp DNA linear HTG 18-FEB-2001
DEFINITION Homo sapiens chromosome 2 clone RP11-550023, WORKING DRAFT
SEQUENCE, 9 unordered pieces.
ACCESSION AC013732
VERSION AC013732.6 GI:9838265
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 141605)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 141605)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (13-NOV-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Aug 17, 2000 this sequence version replaced gi:8961220.

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0550023
----- Summary Statistics -----
Sequencing vector: M13; 73%
Sequencing vector: plasmid; 27%
Chemistry: Dye-primer ET; 73% of reads
Chemistry: Dye-terminator Big Dye; 27% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 136598 bases at least Q40
Consensus quality: 138339 bases at least Q30
Consensus quality: 139218 bases at least Q20
Insert size: 144000; agarose-fp
Insert size: 141800; sum-of-contigs
Quality coverage: 4.32 in Q20 bases; agarose-fp
Quality coverage: 4.85 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 5272: contig of 5272 bp in length
* 5273 5372: gap of unknown length
* 5373 9928: contig of 4556 bp in length
* 9929 10028: gap of unknown length
* 10029 19924: contig of 9896 bp in length
* 19925 20024: gap of unknown length
* 20025 33364: contig of 13340 bp in length
* 33365 33464: gap of unknown length
* 33465 47072: contig of 13608 bp in length
* 47073 47172: gap of unknown length
* 47173 67693: contig of 20521 bp in length

```


COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 30, 2000 this sequence version replaced gi:6573937.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L1985
 Center clone name: L15.C.16

----- Summary Statistics
 Sequencing vector: M13; M7815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 157048 bases at least Q40
 Consensus quality: 159564 bases at least Q30
 Consensus quality: 160513 bases at least Q20
 Insert size: 194000; agarose-fp
 Insert size: 161601; sum-of-contigs
 Quality coverage: 4.4 in Q20 bases; agarose-fp
 Quality coverage: 5.3 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently
 * consists of 8 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 * be preserved.

```

1 5072: contig of 5072 bp in length
* 5073 5172: gap of 100 bp
* 5173 12256: contig of 7084 bp in length
* 12257 12356: gap of 100 bp
* 12357 20176: contig of 7820 bp in length
* 20177 20276: gap of 100 bp
* 20277 34761: contig of 14485 bp in length
* 34762 34861: gap of 100 bp
* 34862 63485: contig of 28624 bp in length
* 63486 63585: gap of 100 bp
* 63586 94394: contig of 30809 bp in length
* 94395 94494: gap of 100 bp
* 94495 126560: contig of 32066 bp in length
* 126561 126660: gap of 100 bp
* 126661 162301: contig of 35641 bp in length.

```

FEATURES

```

Source
1..162301
  /organism="Homo.sapiens"
  /db_xref="taxon:9606"
  /clone="RP11-115C16"
  /clone.lib="RPC1-11 Human Male BAC"

misc_feature
1..5072
  /note="assembly_fragment"
  clone_end:SP6
  vector_side:right"
5173..12256
  /note="assembly_fragment"
12357..20176
  /note="assembly_fragment"
20277..34761
  /note="assembly_fragment"
34862..63485
  /note="assembly_fragment"
clone_end:T7
vector_side:right"
63586..94394
  /note="assembly_fragment"
94495..126560
  /note="assembly_fragment"
126661..162301

```

```

/note="assembly_fragment"
BASE COUNT 52447 a 29329 c 29494 g 50328 t 703 others
ORIGIN
Query Match 62.7%; Score 23.2; DB 2; Length 162301;
Best Local Similarity 77.8%; Pred. No. 81;
Matches 28; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 tatattccaccatattccctcccttggtgtggg 37
||||| ||||| ||||| ||||| |||||
Db 112670 TATATTCCACCATATCCACTTCTCTCTACTGTGGG 112635

RESULT 11
LOCUS AP001997 180446 bp DNA linear HTG 30-MAY-2000
DEFINITION Homo sapiens chromosome 11 clone RP11-281D19 map 11q, WORKING DRAFT
SEQUENCE, 15 unordered pieces.
ACCESSION AP001997
VERSION AP001997.2 GI:8117612
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens DNA, clone:RP11-281D19.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 180446)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Vada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens 180,446 genomic DNA of 11q
JOURNAL Published Only in Database (2000) In press
REFERENCE 2 (bases 1 to 180446)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Vada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
DIRECT SUBMISSION
Submitted (08-MAY-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@psc.riken.go.jp,
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)
On May 30, 2000 this sequence version replaced gi:7768922.
----- Genomic Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@psc.riken.go.jp
----- Project Information
Center project name: HumDrafl1
Center clone name: RP11-281D19
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 170936 bases at least Q40
Consensus quality: 175280 bases at least Q30
Consensus quality: 177565 bases at least Q20
Insert size: 179046; sum-of-contigs
Quality coverage: 5.43x in Q20 bases; sum-of-contigs

```

NOTE: This is a 'working draft' sequence. It currently consists of
 15 contigs. The true order of the pieces is not known and their
 order in this sequence record is arbitrary. Gaps between the
 contigs are represented as runs N, but the exact sizes of the gaps
 are unknown. This record will be updated with the finished sequence
 as soon as it is available and the accession number will be
 preserved

```

1 31405 contig of 31405 bp in length
31506 54312 contig of 22807 bp in length
54413 73827 contig of 19415 bp in length
73928 91654 contig of 17727 bp in length
91755 106526 contig of 14772 bp in length
106627 121878 contig of 15252 bp in length

```

121979 132642 contig of 10664 bp in length
 132743 142911 contig of 10169 bp in length
 143012 152327 contig of 9316 bp in length
 152428 159844 contig of 7417 bp in length
 159945 165718 contig of 5774 bp in length
 165819 172001 contig of 6183 bp in length
 172102 175943 contig of 3842 bp in length
 176044 179290 contig of 3247 bp in length
 179391 180446 contig of 1056 bp in length

Sequence updated (26-May-2000).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 15 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 31405: contig of 31405 bp in length
 * 31406 31505: gap of 100 bp
 * 31506 54312: contig of 22807 bp in length
 * 54313 54412: gap of 100 bp
 * 54413 73827: contig of 19415 bp in length
 * 73828 73927: gap of 100 bp
 * 73928 91654: contig of 17727 bp in length
 * 91655 91754: gap of 100 bp
 * 91755 106526: contig of 14772 bp in length
 * 106527 106626: gap of 100 bp
 * 106627 121878: contig of 15252 bp in length
 * 121879 121978: gap of 100 bp
 * 121979 132642: contig of 10664 bp in length
 * 132643 132742: gap of 100 bp
 * 132743 142911: contig of 10169 bp in length
 * 142912 143011: gap of 100 bp
 * 143012 152327: contig of 9316 bp in length
 * 152328 152427: gap of 100 bp
 * 152428 159844: contig of 7417 bp in length
 * 159845 159944: gap of 100 bp
 * 159945 165718: contig of 5774 bp in length
 * 165719 165818: gap of 100 bp
 * 165819 172001: contig of 6183 bp in length
 * 172002 172101: gap of 100 bp
 * 172102 175943: contig of 3842 bp in length
 * 175944 176044: gap of 100 bp
 * 176044 179290: contig of 3247 bp in length
 * 179291 179390: gap of 100 bp
 * 179391 180446: contig of 1056 bp in length.

FEATURES

source
 1. .180446
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="11"
 /map="11q"
 /clone="RP11-281D19"
 misc_feature
 1. .31405
 /note="assembly_fragment"
 misc_feature
 31506. .54312
 /note="assembly_fragment clone_end:sp6 vector_side:left"
 misc_feature
 54413. .73827
 /note="assembly_fragment"
 misc_feature
 73928. .91654
 /note="assembly_fragment"
 misc_feature
 91755. .106526
 /note="assembly_fragment"
 misc_feature
 106627. .121878
 /note="assembly_fragment"
 misc_feature
 121979. .132642
 /note="assembly_fragment"
 misc_feature
 132743. .142911
 /note="assembly_fragment"
 misc_feature
 143012. .152327
 /note="assembly_fragment"
 misc_feature
 152428. .159844

misc_feature
 /note="assembly_fragment"
 159945. .165718
 /note="assembly_fragment"
 misc_feature
 165819. .172001
 /note="assembly_fragment"
 misc_feature
 172102. .175943
 /note="assembly_fragment clone_end:T7 vector_side:left"
 misc_feature
 176044. .179290
 /note="assembly_fragment"
 misc_feature
 179391. .180446
 /note="assembly_fragment"
 BASE COUNT 58765 a 32357 c 31761 g 56162 t 1401 others
 ORIGIN

Query Match 62.7%; Score 23.2; DB 2; Length 180446;
 Best Local Similarity 77.8%; Pred. No. 79;
 Matches 28; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Oy 2 tatatccaccatcccccttcccccttcgtgtggtgg 37
 ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 17658 TATATTCACCAATATCCACTTCCTCTACTGTGGG 17693

RESULT 12

AC094327 128481 bp DNA linear HTG 20-DEC-2001
 LOCUS Rattus norvegicus clone CH230-3123, *** SEQUENCING IN PROGRESS ***,
 DEFINITION 67 unordered pieces.
 AC094327
 VERSION AC094327.2 GI:17941045
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE

1 (bases 1 to 128481)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
 Benton,J., Blinaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
 Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
 Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
 Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
 Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
 Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
 Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
 Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
 Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
 Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
 Hollins,B., Honsi,E., Howard,S., Huber,J., Hulyk,S., Hume,J.,
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
 Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
 Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
 Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W.,
 Louissegh,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
 Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
 Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
 Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
 Ogulu,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
 Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
 Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
 Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N.,
 Slisnon,I., Sodergren,E., Sonake,T., Sparks,A., Stanley,H.,
 Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
 Tang,H., Tansey,J., Taylor,C., Taylor,T., Tellrod,B., Thomas,N.,
 Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalona,D., Vinson,R.,

Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
 Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S.,
 Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G. and Gibbs, R.

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

Direct Submission
 Unpublished
 2 (bases 1 to 128481)
 Worley, K.C.

Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Dec 20, 2001 this sequence version replaced gi:15624160.

COMMENT

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GAJW
 Center clone name: CH230-3123
 ----- Summary Statistics

Assembly program: Phrap; version 0.990329First call to

findphraplist
 Consensus quality: 108264 bases at least Q40
 Consensus quality: 116155 bases at least Q30
 Consensus quality: 122636 bases at least Q20
 Estimated insert size: 93674; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-fp estimation
 Quality coverage: 1.1x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 67 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 5834: contig of 5834 bp in length
 5835 5934: gap of unknown length
 5935 9357: contig of 3423 bp in length
 9358 9458: gap of unknown length
 9459 12732: contig of 3275 bp in length
 12733 12832: gap of unknown length
 12833 17592: contig of 4860 bp in length
 17593 17792: gap of unknown length
 17793 20127: contig of 2334 bp in length
 20127 20226: gap of unknown length
 20227 23758: contig of 3532 bp in length
 23759 23858: gap of unknown length
 23859 26574: contig of 2716 bp in length
 26575 26674: gap of unknown length
 26675 28571: contig of 1897 bp in length
 28572 28671: gap of unknown length
 28672 30715: contig of 2044 bp in length
 30716 30815: gap of unknown length
 30816 33288: contig of 2473 bp in length
 33289 33388: gap of unknown length
 33389 35435: contig of 2047 bp in length
 35436 35535: gap of unknown length
 35536 37858: contig of 2323 bp in length
 37859 37958: gap of unknown length
 40008: contig of 2050 bp in length
 40009 40108: gap of unknown length
 40109 42562: contig of 2454 bp in length
 42563 42662: gap of unknown length
 42663 44720: contig of 2058 bp in length
 44721 44820: gap of unknown length
 44821 46542: contig of 1722 bp in length
 46543 46642: gap of unknown length

46643 48842: contig of 2200 bp in length
 48843 48942: gap of unknown length
 48943 50541: contig of 1599 bp in length
 50542 50641: gap of unknown length
 50642 52001: contig of 1360 bp in length
 52002 52101: gap of unknown length
 52102 54794: contig of 2693 bp in length
 54795 54895: gap of unknown length
 54896 57037: contig of 2143 bp in length
 57038 57137: gap of unknown length
 57138 59117: contig of 2180 bp in length
 59117 60609: gap of unknown length
 60610 60709: contig of 1192 bp in length
 60710 61964: contig of 1254 bp in length
 61964 62064: gap of unknown length
 62064 63265: contig of 1202 bp in length
 63266 63366: gap of unknown length
 63366 64804: contig of 1438 bp in length
 64804 66730: contig of 1827 bp in length
 66731 66830: gap of unknown length
 66831 68940: contig of 2110 bp in length
 68941 69041: gap of unknown length
 69041 70248: contig of 1208 bp in length
 70249 70348: gap of unknown length
 70349 72230: contig of 1882 bp in length
 72231 72330: gap of unknown length
 72331 74080: contig of 1750 bp in length
 74081 74181: gap of unknown length
 74181 75357: contig of 1176 bp in length
 75357 76335: gap of unknown length
 76336 76935: contig of 1379 bp in length
 76936 79109: contig of 2174 bp in length
 79110 79209: gap of unknown length
 79210 80761: contig of 1552 bp in length
 80762 80862: gap of unknown length
 80862 81868: contig of 1006 bp in length
 81868 83544: contig of 1577 bp in length
 83545 83645: gap of unknown length
 83645 85268: contig of 1623 bp in length
 85268 86522: contig of 1155 bp in length
 86523 86622: gap of unknown length
 86623 88378: contig of 1756 bp in length
 88379 88478: gap of unknown length
 88479 89686: contig of 1208 bp in length
 89687 91911: contig of 2125 bp in length
 91912 92011: gap of unknown length
 92012 93804: contig of 1793 bp in length
 93805 93904: gap of unknown length
 93905 95607: contig of 1703 bp in length
 95608 96978: gap of unknown length
 96979 97078: contig of 1271 bp in length
 97079 98824: contig of 1746 bp in length
 98825 98925: gap of unknown length
 98925 100341: contig of 1417 bp in length
 100342 100441: gap of unknown length
 100442 101607: contig of 1166 bp in length
 101608 101707: gap of unknown length
 101708 103210: contig of 1503 bp in length
 103211 103310: gap of unknown length
 103311 104908: contig of 1598 bp in length
 104909 105009: gap of unknown length
 105009 106191: contig of 1083 bp in length
 106192 106992: gap of unknown length
 106992 107577: contig of 1386 bp in length
 107578 107678: gap of unknown length
 107678 108700: contig of 1023 bp in length

repeat_region 10849. .11144
/note="AluSp repeat: matches 1. .296 of consensus"
repeat_region 11145. .11380
/note="L1ME3 repeat: matches 5884. .6122 of consensus"
repeat_region 11565. .11727
/note="FAM repeat: matches 1. .163 of consensus"
repeat_region 11728. .11805
/note="39 copies 2 mer ta 93% conserved"
repeat_region 11868. .12182
/note="AluSc repeat: matches 1. .309 of consensus"
repeat_region 12362. .12677
/note="L2 repeat: matches 2342. .2681 of consensus"
repeat_region 13235. .13530
/note="AluSc repeat: matches 1. .295 of consensus"
repeat_region 13531. .13782
/note="AluX repeat: matches 47. .298 of consensus"
repeat_region 13799. .13885
/note="AluSc repeat: matches 222. .308 of consensus"
repeat_region 13944. .14062
/note="7SLRNA repeat: matches 236. .317 of consensus"
repeat_region 14066. .14138
/note="L2 repeat: matches 2161. .2240 of consensus"
repeat_region 14163. .14310
/note="MIR repeat: matches 38. .204 of consensus"
repeat_region 14349. .14630
/note="AluJo repeat: matches 1. .282 of consensus"
repeat_region 14671. .14852
/note="L2 repeat: matches 2551. .2743 of consensus"
repeat_region 14965. .15097
/note="FLAM.C repeat: matches 1. .133 of consensus"
repeat_region 15117. .15325
/note="MSTD repeat: matches 193. .394 of consensus"
repeat_region 15432. .15483
/note="26 copies 2 mer tc 80% conserved"
repeat_region 15486. .15659
/note="AluJo repeat: matches 123. .299 of consensus"
repeat_region 15660. .15968
/note="AluSg repeat: matches 1. .309 of consensus"
repeat_region 15974. .16005
/note="16 copies 2 mer tt 100% conserved"
repeat_region 16030. .16167
/note="AluJo repeat: matches 1. .138 of consensus"
repeat_region 16209. .16519
/note="AluJo repeat: matches 3. .310 of consensus"
repeat_region 16540. .16923
/note="MER41D repeat: matches 1. .435 of consensus"
repeat_region 16924. .17238
/note="AluSg repeat: matches 1. .313 of consensus"
repeat_region 17239. .17360
/note="MER41D repeat: matches 435. .557 of consensus"
repeat_region 17533. .17844
/note="AluSg repeat: matches 1. .312 of consensus"
repeat_region 17872. .18007
/note="AluJo repeat: matches 1. .137 of consensus"
repeat_region 18008. .18298
/note="AluSg repeat: matches 1. .291 of consensus"
repeat_region 18299. .18471
/note="AluJo repeat: matches 137. .306 of consensus"
repeat_region 18605. .18762
/note="AluSg/x repeat: matches 135. .291 of consensus"
repeat_region 18763. .19071
/note="AluSg repeat: matches 1. .310 of consensus"
repeat_region 19072. .19102
/note="AluSg/x repeat: matches 105. .135 of consensus"
repeat_region 19104. .19416
/note="AluX repeat: matches 1. .312 of consensus"
repeat_region 19551. .19861
/note="AluJo repeat: matches 1. .311 of consensus"
repeat_region 20477. .20516
/note="MER4B repeat: matches 500. .540 of consensus"
repeat_region 20541. .20840
/note="AluX repeat: matches 1. .301 of consensus"
repeat_region 22154. .22344

Query Match 61.1%; Score 22.6; DB 9; Length 163338;
Best Local Similarity 75.7%; Pred. No. 1.3e+02;
Matches 28; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 ctatataccaccatattcccttcccttcggtgtggg 37
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 21486 CHTGTTCACCCACATCCCTCTCTCTA|CCTCTGGG 21450
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14
AC007430/c
LOCUS AC007430
DEFINITION Homo sapiens chromosome 9 clone RP11-94E8 map 9, *** SEQUENCING IN
PROGRESS ***, 7 unordered pieces.
AC007430
AC007430
AC007430.18 GI:8705107
HTG: HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
VERSION
KEYWORDS
SOURCE
ORGANISM

Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 172588)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 9, clone RP11-94E8
Unpublished
2 (bases 1 to 172588)

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L696
Center clone name: 94_E_8

NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1
* 2062 2161: gap of 100 bp
* 2162 3140: contig of 979 bp in length
* 3141 3240: gap of 100 bp
* 3241 6736: contig of 3496 bp in length
* 6737 6836: gap of 100 bp
* 6837 120051: contig of 113215 bp in length

* 120052 120151: gap of 100 bp
 * 120152 122648: contig of 2497 bp in length
 * 122649 122748: gap of 100 bp
 * 122749 125082: contig of 2334 bp in length
 * 125083 125182: gap of 100 bp
 * 125183 172588: contig of 47406 bp in length.
 FEATURES Location/Qualifiers
 source 1..172588
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="9"
 /map="9"
 /clone="RP11-94E8"
 /clone_lib="RPC1-11 Human Male BAC"
 BASE COUNT 41981 a 43766 c 42666 g 43574 t 601 others
 ORIGIN

Query Match 61.18; Score 22.6; DB 2; Length 172588;
 Best Local Similarity 75.7%; Pred. No. 1.3e+02;
 Matches 28; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 ctatcatcaccatccctccctccctcgctggtgg 37
 Db 9187 CTGTGTTACCCACATCCCTCTCTAGCTCTGGG 9151

RESULT 15
 AC093959/c
 LOCUS AC093959 236135 bp DNA linear HTG 20-DEC-2001
 DEFINITION Rattus norvegicus clone CH230-3B9, *** SEQUENCING IN PROGRESS ***,
 51 unordered pieces.
 AC093959
 VERSION AC093959.3 GI:17940888
 KEYWORDS HTG; HTGS PHASE1.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

1 (bases 1 to 236135)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
 Benton,J., Blmage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
 Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
 Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
 Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
 Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
 Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
 Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
 Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
 Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
 Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
 Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
 Hollins,B., Honsi,J., Howard,S., Huber,J., Hulyk,S., Hume,J.,
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
 Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
 Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
 Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Lucier,R., Liu,W.,
 Louisaged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
 Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
 Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,
 Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
 Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S.,
 Ogulu,M., Okuwonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
 Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
 Quiles,M., Ren,Y., Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M.,
 Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
 Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,

Stone,E., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
 Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
 Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
 Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
 Watlington,S., Williams,G., Williamson,A., Wlarczyk,R., Woodson,S.,
 Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstein,G., and Gibbs,R.
 Direct Submission
 Unpublished
 2 (bases 1 to 236135)
 Worley,K.C.
 Direct Submission
 Submitted (13-SEP-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Dec 20, 2001 this sequence version replaced gi:15627002.

Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GADK
 Center clone name: CH230-3B9
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329First call to
 findPhrapList

Consensus quality: 200925 bases at least Q40
 Consensus quality: 210235 bases at least Q30
 Consensus quality: 217064 bases at least Q20
 Estimated insert size: 217481; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-fp estimation
 Quality coverage: 2.9x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 51 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 21130: contig of 21130 bp in length
 * 21131 21230: gap of unknown length
 * 21231 39576: contig of 18346 bp in length
 * 39577 39676: gap of unknown length
 * 39677 52840: contig of 13164 bp in length
 * 52841 52940: gap of unknown length
 * 52941 64146: contig of 11206 bp in length
 * 64147 73196: gap of unknown length
 * 73197 73296: gap of unknown length
 * 73297 82347: contig of 9051 bp in length
 * 82348 88373: gap of unknown length
 * 88374 88473: gap of unknown length
 * 88474 96416: contig of 7943 bp in length
 * 96417 96516: gap of unknown length
 * 96517 102864: contig of 6348 bp in length
 * 102865 102964: gap of unknown length
 * 102965 109031: contig of 6067 bp in length
 * 109032 109131: gap of unknown length
 * 109132 117283: contig of 8152 bp in length
 * 117284 117383: gap of unknown length
 * 117384 122439: contig of 5056 bp in length
 * 122440 122539: gap of unknown length
 * 122540 129255: contig of 6716 bp in length
 * 129256 129355: gap of unknown length
 * 129356 133291: contig of 3936 bp in length
 * 133292 133391: gap of unknown length
 * 133392 138423: contig of 5038 bp in length

THIS PAGE BLANK (USPTO)

Result No.	Score	Query %			DB	ID	Description
		Match	Length				
C	1	24.2	60.5	288	9	AI786367	AI786367 uj54805.x
	2	24	60.0	528	9	AI751810	AI751810 AU151810
	3	23.4	58.5	754	9	AA811417	AA811417 OB82H01.s
C	4	23	57.5	378	9	BE142441	BE142441 RC2-HT014
	5	23	57.5	465	10	BI014164	BI014164 PM3-ET020
	6	23	57.5	606	10	BE531296	BE531296 G01278222
C	7	22.6	56.5	753	10	BG774850	BG774850 602649944
	8	22.6	56.5	929	10	BM453604	BM453604 AGENCEOURT
	9	22.4	56.0	281	10	BI063933	BI063933 IL3-UT011
C	10	22.4	56.0	1064	10	BE260926	BE260926 G01153879
	11	22.2	55.5	369	10	R25470	R25470 yH41905.r1
	12	22.2	55.5	441	10	BE490643	BE490643 WHE0370.B
C	13	22.2	55.5	464	10	BE444887	BE444887 WHE1129_D
	14	22.2	55.5	489	10	BE516924	BE516924 WHE0622_H
	15	22.2	55.5	616	10	BG607929	BG607929 WHE2475_D
C	16	22.2	55.5	629	10	BE585665	BE585665 EST46PSP6
	17	22.2	55.5	1169	10	BG3411375	BG3411375 602463854

Qy 6 tccaccatatcccccttgggcccttggggtgtg 38

Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LCM266 row: k column: 05
 High quality sequence stop: 404.
 Location/Qualifiers

FEATURES

source
 1..606
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3610492"
 /clone_lib="NIH_MGC_39"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
 Site_2: EcoRI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCAGGAG(G). Library constructed
 by Ling Hong in the laboratory of Gerald M. Rubin
 (University of California, Berkeley) using ZAP-cDNA
 synthesis kit (Stratagene) and Superscript II RT (Life
 Technologies)."
 BASE COUNT 129 a 172 c 175 g 128 t 2 others
 ORIGIN

Query Match 57.5%; Score 23; DB 10; Length 606;
 Best Local Similarity 74.4%; Pred. No. 2.3e+02;
 Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 tatttccaccatcccccttggcccttgggtgtgg 40
 | ||||| ||||| | ||||| ||||| ||||| |||||
 Db 539 TTTTTCCTCCCAAGCCCAAGCGCTTGGGTTTGGG 577

RESULT 7
 BG774850 753 bp mRNA linear EST 15-MAY-2001
 LOCUS 602649944F1 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:4760867 5',
 DEFINITION mRNA sequence.
 ACCESSION BG774850
 VERSION BG774850.1 GI:14045167
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 753)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: DCTD/DTF

cdna Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LCM1612 row: g column: 12
 High quality sequence stop: 435.
 Location/Qualifiers

FEATURES

source
 1..753
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4760867"
 /clone_lib="NIH_MGC_40"
 /tissue_type="carcinoma, cell line"

/lab_host="DH10B (phage-resistant)"
 /note="Organ: prostate; Vector: pOTB7; Site_1: XhoI;
 Site_2: EcoRI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCAGGAG(G). Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC Library."
 BASE COUNT 232 a 163 c 174 g 184 t
 ORIGIN

Query Match 56.5%; Score 22.6; DB 10; Length 753;
 Best Local Similarity 75.7%; Pred. No. 3.4e+02;
 Matches 28; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 ttctccaccatcccccttggcccttgggtgtgg 40
 | ||||| ||||| | ||||| ||||| ||||| |||||
 Db 541 TTTCCACACATATGCACACTGGCCCTCGCATGTGG 577

RESULT 8
 BM453604/c 929 bp mRNA linear EST 05-FEB-2002
 LOCUS AGENCOURT_6401106 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5493739
 DEFINITION 5', mRNA sequence.
 ACCESSION BM453604
 VERSION BM453604.1 GI:18502644
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 929)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC

cdna Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM12117 row: o column: 20
 High quality sequence start: 180
 High quality sequence stop: 404.
 Location/Qualifiers

FEATURES
 source
 1..929
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5493739"
 /clone_lib="NIH_MGC_67"
 /tissue_type="retinoblastoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.75 kb. Library constructed by Life
 Technologies."

BASE COUNT 273 a 228 c 248 g 178 t 2 others
 ORIGIN

Query Match 56.5%; Score 22.6; DB 10; Length 929;
 Best Local Similarity 75.7%; Pred. No. 3.5e+02;
 Matches 28; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 3 atttccaccatcccccttggcccttgggtgtgg 39
 | ||||| ||||| | ||||| ||||| ||||| |||||
 Db 586 ATTTTCCCTCCATTTTCTTTGGGCCCTTAGGGGGG 550


```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1064)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-femail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCMI93 row: n column: 20
High quality sequence stop: 735.
FEATURES
Location/Qualifiers
1..1064
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3509971"
/clone_lib="NIH_MGC_19"
/tissue_type="neuroblastoma"
/lab_host="DH10B {phage-resistant}"
/note="Organ: brain; Vector: pOT87; Site_1: xhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 230 a 345 c 277 g 212 t
ORIGIN
Query Match 56.0%; Score: 22.4; DB 10; Length 1064;
Best Local Similarity 81.2%; Pred. No. 4.2e+02;
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 3 attccaccaccatccccctggcccttggg 34
| | | | | | | | | | | | | | | | | |
Db 885 ACTTCCACAAATATCCCTTGGCGCGGTG 916
RESULT 11
R25470/c
LOCUS R25470
DEFINITION YH41905.r1 Soares placenta Nb2HP Homo sapiens CDNA clone
IMAGE:132344 5', mRNA sequence.
ACCESSION R25470
VERSION R25470.1 GI:791605
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 369)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Maria,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevasakis,E., Waterston
,R., Williamson,A., Wohlmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1225
High quality sequence stops: 239

```

Source: IMAGE Consortium, LLNL
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 1225 Std Error: 0.00
 Seq primer: M13Rpl
 High quality sequence stop: 239.
 Location/Qualifiers

FEATURES

source

1. .369
 /organism="Homo sapiens"
 /db_xref="GB:537936"
 /db_xref="taxon:9606"
 /clone="IMAGE:13234"
 /clone_lib="Soares placenta Nb2HP"
 /sex="Female"
 /dev_stage="placenta obtained at birth (full term)"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a
 modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
 strand cDNA was primed with a Not I - Oligo(dT) primer [5'
 AACTGGGAATTCGGCGCCGAGGAATTTTTTTTTTTTTTTT 3'],
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT7T3 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M.Fatima Bonaldo.
 103 a 58 c 109 g 97 t 2 others

BASE COUNT

ORIGIN

Query Match 55.5%; Score 22.2; DB 10; Length 369;
 Best Local Similarity 77.1%; Pred. No. 4.2e+02;
 Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ctattccaccaccatcccttggcccttgggt 35
 ||| ||||| ||| ||| ||||| ||||| |||||
 Db 306 CTACTTCCTTCCTTCCTCCTTGGCTTGGGT 272

RESULT 12

BE490643

LOCUS
 DEFINITION WHE0370_B03_D06ZS Wheat cold-stressed seedling cDNA library
 Triticum aestivum cDNA clone WHE0370_B03_D06, mRNA sequence.

ACCESSION BE490643
 VERSION BE490643.1 GI:9610176

KEYWORDS
 SOURCE EST.
 bread wheat.

ORGANISM

Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
 ; Triticeae; Triticum.

REFERENCE

AUTHORS

Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han
 P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,
 Seaton,C.L. and Tong,J.C.
 The structure and function of the expressed portion of the wheat
 genomes - Cold-stressed seedling cDNA library
 Unpublished (2000)

JOURNAL

COMMENT

Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific
 West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA

Tel: 5105595773
 Fax: 5105595818

Email: oanderson@pw.usda.gov
 Sequence have been trimmed to remove vector sequence and low
 quality sequence with phred score less than 20

Seq primer: Stratagene SK primer.
 Location/Qualifiers

FEATURES

source

1. .441
 /organism="Triticum aestivum"
 /cultivar="Chinese Spring"
 /db_xref="taxon:4565"

/clone="WHE0370_B03_D06"
 /clone_lib="Wheat cold-stressed seedling cDNA library"
 /tissue_type="Seedling"
 /dev_stage="Five-day old seedling"
 /lab_host="E. coli SOLR"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
 Site_1: ECORI; Site_2: XhoI; Seeds were surface-sterilized
 , germinated and grown aseptically in the dark at room
 temperature on filter paper with water, nystatin and
 cefotaxime in covered crystallization dishes. Five-day
 old seedlings were transferred to 5 C cold room and kept
 for 48 hr. The tissue, total RNA, and poly(A) RNA were
 prepared, a cDNA library was made, and the cDNA clones
 were in vivo excised to give pluescript phagemids in the
 TJ Close lab (Choi, Close, Fenton) at the University of
 California, Riverside. Plasmid DNA preparations and DNA
 sequencing were performed in the OD Anderson lab (all
 other authors)."
 84 a 121 c 102 g 134 t

BASE COUNT

ORIGIN

Query Match 55.5%; Score 22.2; DB 10; Length 441;
 Best Local Similarity 77.1%; Pred. No. 4.3e+02;
 Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 4 ttccccaccatcccttggcccttgggtgtg 38
 ||| ||||| ||| ||| ||||| ||||| |||||
 Db 246 TTACCCCACTTTCCTTGGCTTGGGTGTG 280

RESULT 13

BE444887

LOCUS

DEFINITION WHE1129_D06_G11ZS wheat etiolated seedling root normalized cDNA
 library Triticum aestivum cDNA clone WHE1129_D06_G11, mRNA
 sequence.

ACCESSION BE444887
 VERSION BE444887.1 GI:9444439

KEYWORDS

SOURCE

EST.
 bread wheat.

ORGANISM

Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
 ; Triticeae; Triticum.

REFERENCE

AUTHORS

Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han
 P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Nguyen,H.T.,
 Rausch,C.J., Seaton,C.L., Tong,J.C. and Zhang,D.
 The structure and function of the expressed portion of the wheat
 genomes - Normalized root cDNA library
 Unpublished (2000)

JOURNAL

COMMENT

Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific
 West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA

Tel: 5105595773
 Fax: 5105595818

Email: oanderson@pw.usda.gov
 Sequence have been trimmed to remove vector sequence and low
 quality sequence with phred score less than 20

Seq primer: Stratagene SK primer.
 Location/Qualifiers

1. .464
 /organism="Triticum aestivum"
 /cultivar="Chinese Spring"
 /db_xref="taxon:4565"

/clone="WHE1129_D06_G11"
 /clone_lib="Wheat etiolated seedling root normalized cDNA
 library"
 /tissue_type="Root"
 /dev_stage="Five day old etiolated seedling"
 /lab_host="E. coli DH10B"

RNA were prepared by Steven Verhey in M.K. Walker-Simmons's lab (USDA-ARS, Washington State Univ., Pullman, Washington 99164-6420). A cDNA library was made by Clontech using a combination of random and oligo dt primers. Library was plated and archived by Russell Johnson (Colby College, ME/Walker-Simmons' lab). Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors).*

```
Query Match      55.5%; Score: 22.2; DB 10; Length 489;
Best Local Similarity 77.1%; Pred. NO. 4.e+02;
Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
```

RESULT	15
LOCUS	BG607929
DEFINITION	WHE2475_D11_G2L2S Triticum monococcum early reproductive apex cDNA library Triticum monococcum CDNA clone WHE2475_D11_G2L mRNA sequence.
	616 bp mRNA linear EST 17-APR-2001

ACCESSION	B6607929
VERSION	B6607929.1
KEYWORDS	GI:13657912
SOURCE	EST.
ORGANISM	Triticum monococcum
REFERENCE	Triticum monococcum
AUTHORS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae ; Triticeae; Triticum.
TITLE	1 (bases 1 to 616)
	Anderson, O.D., Chao, S., Dubcovsky, J., Echenique, V., Han, P. S., Hsia C.-C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J., Seaton, C.L., Stamova, B. and Tong, J.C.
	The structure and function of the expressed portion of the wheat genomes - Early reproductive apex cDNA library from Triticum monococcum

JOURNAL
COMMENT

Unpublished (2001)
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 510595773
Fax: 510595818
Email: candersn@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: Stratagene SK primer.

```

/clone_lib="WHE2475_D11_G21"
/clone_lib="Triticum monococcum early reproductive apex
cDNA library"
/tissue_type="Early reproductive apex"
/dev_stage="Seven week-old plants"
/lab_host="E. coli XL0LR"
/notes="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site_1: EcoRI; Site_2: XhoI; The tissue, total RNA, and
poly(A) RNA were prepared from apex at double-ridge stage
to terminal-spikelat stage during transition from
vegetative state to flower state, a cDNA library was made,
and the cDNA clones were in vivo excised at the
University of California, Davis (V. Echenique, B. Stamova
J. Dubcovsky). Plasmid DNA preparations and DNA
sequencing were performed in the OD Anderson lab (all

```


Db 805 CTGTCACACATCTCTCTCAGCAGCTCGGTGGTGG 844

RESULT 5

US-08-878-989-13

; Sequence 13, Application US/08878989

; Patent No. 5885803

; GENERAL INFORMATION:

; APPLICANT: Bandman, Olga

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Corley, Neil C.

; APPLICANT: Guegler, Karl G.

; APPLICANT: Lal, Preeti

; APPLICANT: Goli, Surya K.

; APPLICANT: Shah, Purvi

; TITLE OF INVENTION: DISEASE ASSOCIATED PROTEIN

; TITLE OF INVENTION: KINASES

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; APPLICATION NUMBER: US/08/878,989

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J J

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0321 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-855-0555

; TELEFAX: 415-845-4166

; TELEX:

; INFORMATION FOR SEQ ID NO: 13:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1866 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: COLNUT03

; CLONE: 1340712

; US-08-878-989-13

Query Match 47.5%; Score 19; DB 2; Length 1866;

Best Local Similarity 71.4%; Pred. No. 50;

Matches 25; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 5 ttccaccaccatcccccttggtggccttggtggtg 39

| | | | | | | | | | | | | | | | | | | | |

Db 165 TACCACCCACACCCCTTGCCCATTTTGGTGC 199

RESULT 6

US-09-272-796-13

; Sequence 13, Application US/09272796

; Patent No. 6207148

; GENERAL INFORMATION:

; APPLICANT: Bandman, Olga

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Corley, Neil C.

; APPLICANT: Guegler, Karl G.

; APPLICANT: Lal, Preeti

; APPLICANT: Goli, Surya K.

; APPLICANT: Shah, Purvi

; TITLE OF INVENTION: DISEASE ASSOCIATED PROTEIN

; TITLE OF INVENTION: KINASES

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/272,796

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/878,989

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J J

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0321 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-855-0555

; TELEFAX: 415-845-4166

; TELEX:

; INFORMATION FOR SEQ ID NO: 13:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1866 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: COLNUT03

; CLONE: 1340712

; US-09-272-796-13

Query Match 47.5%; Score 19; DB 4; Length 1866;

Best Local Similarity 71.4%; Pred. No. 50;

Matches 25; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 5 ttccaccaccatcccccttggtggccttggtggtg 39

| | | | | | | | | | | | | | | | | | | | |

Db 165 TACCACCCACACCCCTTGCCCATTTTGGTGC 199

RESULT 7

US-08-665-259-24/c

; Sequence 24, Application US/08665259

; Patent No. 6028173

; GENERAL INFORMATION:

; APPLICANT: Landes, Gregory M.

; APPLICANT: Burn, Timothy C.

; APPLICANT: Connors, Timothy D.

; APPLICANT: Dackowski, William R.

; APPLICANT: Van Raay, Terence J.

; APPLICANT: Klinger, Katherine W.

; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,

; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME

; NUMBER OF SEQUENCES: 73

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: GENZYME CORPORATION

; STREET: One Mountain Road

:/ CITY: Framingham
:/ STATE: Massachusetts
:/ COUNTRY: United States of America
:/ ZIP: 01701
:/ COMPUTER READABLE FORM:
:/ MEDIUM TYPE: Floppy disk
:/ COMPUTER: IBM PC compatible
:/ OPERATING SYSTEM: PC-DOS/MS-DOS
:/ SOFTWARE: PatentIn Release #1.0, Version #1.30
:/ CURRENT APPLICATION DATA:
:/ APPLICATION NUMBER: US/08/665,259
:/ FILING DATE: 17-JUN-1996
:/ CLASSIFICATION: 435
:/ ATTORNEY/AGENT INFORMATION:
:/ NAME: Dugan, Deborah A.
:/ REGISTRATION NUMBER: 37,315
:/ REFERENCE/DOCKET NUMBER: IG5-9.1
:/ TELECOMMUNICATION INFORMATION:
:/ TELEPHONE: (508) 872-8400
:/ TELEFAX: (508) 872-5415
:/ INFORMATION FOR SEQ ID NO: 24:
:/ SEQUENCE CHARACTERISTICS:
:/ LENGTH: 5894 base pairs
:/ TYPE: nucleic acid
:/ STRANDEDNESS: single
:/ TOPOLOGY: linear
:/ MOLECULE TYPE: cDNA
:/ FEATURE:
:/ NAME/KEY: CDS
:/ LOCATION: 2..5053
:/ US-08-665-259-24

Query Match 47.5%; Score 19; DB 3; Length 5894;
Best Local Similarity 71.4%; Pred. No. 59;
Matches 25; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 tatttccaccatccccctgggcccctgggtg 36
||||| | | | | | | | | | | | | | | | | | | |
Db 1571 TATTTCACCCCTGAACACCTTGGACAGGTGCTTG 1537

RESULT 8
US-08-762-500-24/c
:/ Sequence 24, Application US/08762500
:/ Patent No. 6030806
:/ GENERAL INFORMATION:
:/ APPLICANT: Landes, Gregory M.
:/ APPLICANT: Burn, Timothy C.
:/ APPLICANT: Connors, Timothy D.
:/ APPLICANT: Dackowski, William R.
:/ APPLICANT: Van Raay, Terence J.
:/ APPLICANT: Klinger, Katherine W.
:/ TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
:/ NUMBER OF SEQUENCES: 83
:/ CORRESPONDENCE ADDRESS:
:/ ADDRESSEE: GENZYME CORPORATION
:/ STREET: One Mountain Road
:/ CITY: Framingham
:/ STATE: Massachusetts
:/ COUNTRY: United States of America
:/ ZIP: 01701
:/ COMPUTER READABLE FORM:
:/ MEDIUM TYPE: Floppy disk
:/ COMPUTER: IBM PC compatible
:/ OPERATING SYSTEM: PC-DOS/MS-DOS
:/ SOFTWARE: PatentIn Release #1.0, Version #1.30
:/ CURRENT APPLICATION DATA:
:/ APPLICATION NUMBER: US/08/762,500
:/ FILING DATE: 09-DEC-1996
:/ CLASSIFICATION: 435
:/ PRIOR APPLICATION DATA:

:/ APPLICATION NUMBER: US 08/665,259
:/ FILING DATE: 17-JUN-1996
:/ PRIOR APPLICATION DATA: PCT/US96/10469
:/ APPLICATION NUMBER: PCT/US96/10469
:/ FILING DATE: 17-JUN-1996
:/ ATTORNEY/AGENT INFORMATION:
:/ NAME: Dugan, Deborah A.
:/ REGISTRATION NUMBER: 37,315
:/ REFERENCE/DOCKET NUMBER: IG5-9.3
:/ TELECOMMUNICATION INFORMATION:
:/ TELEPHONE: (508) 872-8400
:/ TELEFAX: (508) 872-5415
:/ INFORMATION FOR SEQ ID NO: 24:
:/ SEQUENCE CHARACTERISTICS:
:/ LENGTH: 5894 base pairs
:/ TYPE: nucleic acid
:/ STRANDEDNESS: single
:/ TOPOLOGY: linear
:/ MOLECULE TYPE: cDNA
:/ FEATURE:
:/ NAME/KEY: CDS
:/ LOCATION: 2..5053
:/ US-08-762-500-24

Query Match 47.5%; Score 19; DB 3; Length 5894;
Best Local Similarity 71.4%; Pred. No. 59;
Matches 25; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 tatttccaccatccccctgggcccctgggtg 36
||||| | | | | | | | | | | | | | | | | | | |
Db 1571 TATTTCACCCCTGAACACCTTGGACAGGTGCTTG 1537

RESULT 9
US-08-762-500-74/c
:/ Sequence 74, Application US/08762500
:/ Patent No. 6030806
:/ GENERAL INFORMATION:
:/ APPLICANT: Landes, Gregory M.
:/ APPLICANT: Burn, Timothy C.
:/ APPLICANT: Connors, Timothy D.
:/ APPLICANT: Dackowski, William R.
:/ APPLICANT: Van Raay, Terence J.
:/ APPLICANT: Klinger, Katherine W.
:/ TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
:/ NUMBER OF SEQUENCES: 83
:/ CORRESPONDENCE ADDRESS:
:/ ADDRESSEE: GENZYME CORPORATION
:/ STREET: One Mountain Road
:/ CITY: Framingham
:/ STATE: Massachusetts
:/ COUNTRY: United States of America
:/ ZIP: 01701
:/ COMPUTER READABLE FORM:
:/ MEDIUM TYPE: Floppy disk
:/ COMPUTER: IBM PC compatible
:/ OPERATING SYSTEM: PC-DOS/MS-DOS
:/ SOFTWARE: PatentIn Release #1.0, Version #1.30
:/ CURRENT APPLICATION DATA:
:/ APPLICATION NUMBER: US/08/762,500
:/ FILING DATE: 09-DEC-1996
:/ CLASSIFICATION: 435
:/ PRIOR APPLICATION DATA:
:/ APPLICATION NUMBER: US 08/665,259
:/ FILING DATE: 17-JUN-1996
:/ PRIOR APPLICATION DATA: PCT/US96/10469
:/ APPLICATION NUMBER: PCT/US96/10469
:/ FILING DATE: 17-JUN-1996
:/ ATTORNEY/AGENT INFORMATION:
:/ NAME: Dugan, Deborah A.
:/ REGISTRATION NUMBER: 37,315

REFERENCE/DOCKET NUMBER: IG5-9.3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 872-8400
TELEFAX: (508) 872-5415
INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 6525 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 573..5684
US-08-762-500-74

Query Match 47.5%; Score 19; DB 3; Length 6525;

Best Local Similarity 71.4%; Pred. No. 60; Indels 0; Gaps 0;
Matches 25; Conservative 0; Mismatches 10;

Qy 2 tatttccaccatccccctggcccttggtg 36
Db 2202 TATTTCCACCTGAACACCTTGGACAGGTGCTTG 2168
|||||

RESULT 10

US-08-991-789A-207/c
; Sequence 207, Application US/08991789A
; Patent No. 6225054

GENERAL INFORMATION:

APPLICANT: Frudakis, Tony N.
Smith, John M.
Reed, Steven G.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER

NUMBER OF SEQUENCES: 292

CORRESPONDENCE ADDRESS:

ADDRESSEE: Seed IP Law Group
STREET: 701 Fifth Avenue, Suite 6300
CITY: Seattle
STATE: Washington
COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/991,789A
FILING DATE: 11-Dec-1997
CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Potter, Jane E. R.

REGISTRATION NUMBER: 33,332

REFERENCE/DOCKET NUMBER: 210121.419C3

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 207:

SEQUENCE CHARACTERISTICS:

LENGTH: 176 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 207:

US-08-991-789A-207

Query Match

47.0%; Score 18.8; DB 4; Length 176;

Best Local Similarity 76.7%; Pred. No. 42; Indels 0; Gaps 0;
Matches 23; Conservative 0; Mismatches 7;

Qy 1 ctatttccaccatcccccttgggcct 30
Db 173 CCATTACCACTATCCCATCAGTCACT 144
|||||

RESULT 11

US-09-062-451-207/c

; Sequence 207, Application US/09062451

; Patent No. 6344550

GENERAL INFORMATION:

APPLICANT: Frudakis, Tony N.

APPLICANT: Smith, John M.

APPLICANT: Reed, Steven G.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE

TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF BREAST CANCER

NUMBER OF SEQUENCES: 297

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/062,451

FILING DATE: 04-APR-1997

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.419C2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 207:

SEQUENCE CHARACTERISTICS:

LENGTH: 176 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-09-062-451-207

Query Match

47.0%; Score 18.8; DB 4; Length 176;

Best Local Similarity 76.7%; Pred. No. 42;

Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 ctatttccaccatcccccttgggcct 30

Db 173 CCATTACCACTATCCCATCAGTCACT 144
|||||

RESULT 12

US-09-227-357-41/c

; Sequence 41, Application US/09227357

; Patent No. 6342581

GENERAL INFORMATION:

APPLICANT: Fischer et al.

TITLE OF INVENTION: 123 Human Secreted Proteins

FILE REFERENCE: P2010P1

CURRENT APPLICATION NUMBER: US/09/227,357

CURRENT FILING DATE: 1999-01-08

EARLIER APPLICATION NUMBER: PCT/US98/13684

EARLIER FILING DATE: 1998-07-07

EARLIER APPLICATION NUMBER: 60/051,926

EARLIER FILING DATE: 1997-07-08

EARLIER APPLICATION NUMBER: 60/052,793

EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,925
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,929
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,803
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,732
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,931
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,932
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,916
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,930
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,918
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,920
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,733
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,795
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,919
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,928
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/055,722
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,723
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,948
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,949
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,953
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,950
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,947
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,964
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/056,360
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,684
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,984
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,954
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/058,785
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,664
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,660
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,661
EARLIER FILING DATE: 1997-09-12
NUMBER OF SEQ ID NOS: 672
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 41
LENGTH: 1158
TYPE: DNA
ORGANISM: Homo sapiens
US-09-227-357-41

Query Match 47.0%; Score 18.8; DB 4; Length 1158;
Best Local Similarity 76.7%; Pred. No. 56;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ctatttcacccatattcccttggccct 30
||| ||| ||| ||| ||| ||| ||| |||
Db 49 CTCTTCCCTCCCATTTCCCTCGTGCCAT 20

RESULT 13
US-08-904-032-2/c
; Sequence 2, Application US/08904032
; Patent No. 6004773
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: EMBRYOGENESIS PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA: US/08/904,032
; APPLICATION NUMBER: US/08/904,032
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0359 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2921 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BLADNOT03
; CLONE: 1602555
US-08-904-032-2

Query Match 47.0%; Score 18.8; DB 3; Length 2921;
Best Local Similarity 76.7%; Pred. No. 64;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ctatttcacccatattcccttggccct 30
||| ||| ||| ||| ||| ||| ||| |||
Db 1882 CTGCTTCCACCCCAAGCCCTTGGTTCT 1853

RESULT 14
US-09-134-246-8/c
; Sequence 8, Application US/09134246B
; Patent No. 6207377
; GENERAL INFORMATION:
; APPLICANT: Wayne, Jay
; APPLICANT: Xu, Shuang-yong
; TITLE OF INVENTION: Method For Construction Of Thermus-E. coli Shuttle
; Vectors And Identification Of Two Thermus Plasmid

```

: TITLE OF INVENTION: Replication Origins
: FILE REFERENCE: Thermus Shuttle Vector
: CURRENT APPLICATION NUMBER: US/09/134,246B
: CURRENT FILING DATE: 1998-08-14
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 8
: LENGTH: 11958
: TYPE: DNA
: ORGANISM: Thermus sp.
: US-09-134-246-B

```

Query Match 47.0%; Score 18.8; DB 4; Length 11958;
Best Local Similarity 76.7%; Pred. No. 79;
Matches 23; Conservative 0; Mismatches 7; Indels 0;

Qy 1 ctatttccaccatatacccccttggccct 30
|| ||||| || | ||||| |||||
Db 7859 CTTTTCCTCCCGCCCTTTTCCCCCTGGCCCCCT 7830

RESULT 15

```

US-09-123-851-2/C
; Sequence 2, Application US/09123851
; Patent NO. 5958405
; GENERAL INFORMATION:
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: A NOVEL HUMAN PROTEIN TYROSINE KINASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Inocyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/123,851

```

PRIORITY: 1
 PRIOR APPLICATION DATA: 2
 APPLICATION NUMBER: 08/728,520 3
 FILING DATE: 4
 ATTORNEY/AGENT INFORMATION: 5
 NAME: Billings, Lucy J. 6
 REGISTRATION NUMBER: 36,749 7
 REFERENCE/DOCKET NUMBER: PF-0136 US 8
 TELECOMMUNICATION INFORMATION: 9
 TELEPHONE: 415-855-0555 10
 TELEFAX: 415-845-4166 11
 INFORMATION FOR SEQ ID NO: 2: 12
 SEQUENCE CHARACTERISTICS: 13
 LENGTH: 1346 base pairs 14
 TYPE: nucleic acid 15
 STRANDEDNESS: single 16
 TOPOLOGY: linear 17
 MOLECULE TYPE: cDNA 18
 IMMEDIATE SOURCE: 19
 LIBRARY: 20
 CLONE: Consensus 21
 US-09-123-851-2 22

Query Match 46.0%; Score 18.4; DB 2; Length 1346;
Best Local Similarity 69.4%; Pred. No. 82;
Matches 25; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 4 ttccccccatatacccccttggggcccttgggtgtgg 39

Db 1108 TATGCCCCCGCTTGGCCCCCTGGGCCCCCTGGGCTTG 1073

Search completed: June 4, 2002, 17:47:26
Job time: 6599 sec

THIS PAGE BLANK (USPTO)

corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr1>

RP5-1018K9 is from the library RPCI-5 constructed by the group of Pieter de Jong. For further details see

<http://www.chori.org/bacpac/home.htm>

VECTOR: pCYFAC2

IMPORTANT: This sequence is not the entire insert of clone RP5-1018K9. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true left end of clone RP5-1018K9 is at 1 in this sequence. The true left end of clone RP4-800F24 is at 15401 in this sequence.

FEATURES

```

source
1..125500
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="q23.2-q24.3"
/clone="RP5-1018K9"
/clone_lib="RPCI-5"
1..2074
/note="L1MA4 repeat: matches 4221..6294 of consensus"
2689..3162
/note="L1PA15 repeat: matches 5683..6153 of consensus"
3180..3610
/note="L2 repeat: matches 2035..2495 of consensus"
3677..3785
/note="MER69A repeat: matches 1..110 of consensus"
3792..3956
/note="L2 repeat: matches 1740..1897 of consensus"
3987..4199
/note="L2 repeat: matches 2123..2335 of consensus"
4403..4605
/note="MIR repeat: matches 26..249 of consensus"
4636..4734
/note="L2 repeat: matches 2648..2749 of consensus"
5440..5585
/note="MIR repeat: matches 26..176 of consensus"
5603..5798
/note="MLT1F repeat: matches 352..539 of consensus"
5919..6176
/note="MLT1F repeat: matches 10..252 of consensus"
6388..6637
/note="MIR repeat: matches 3..262 of consensus"
7575..7695
/note="L2 repeat: matches 2623..2744 of consensus"
10912..11284
/note="MLT1A1 repeat: matches 20..365 of consensus"
11990..12450
/note="LTR2 repeat: matches 1..449 of consensus"
12894..13515
/note="match: STS: Em:G56281"
12969..13077
/note="L2 repeat: matches 2575..2690 of consensus"
13338..13551
/note="L2 repeat: matches 2256..2268 of consensus"
13552..13868
/note="AluJo repeat: matches 3..312 of consensus"
13869..14262
/note="L2 repeat: matches 2268..2679 of consensus"

repeat_region
14286..14398
/note="AluSg/x repeat: matches 168..294 of consensus"
15207..15430
/note="56 copies 4 mer cttt 79% conserved"
15225..15432
/note="104 copies 2 mer ct 77% conserved"
15228..15419
/note="6 copies 32 mer 83% conserved"
15232..15434
/note="7 copies 29 mer 66% conserved"
15434..15669
/note="AluJo repeat: matches 75..312 of consensus"
15925..16102
/note="MIR repeat: matches 75..260 of consensus"
16637..16743
/note="L2 repeat: matches 2586..2710 of consensus"
16909..17295
/note="LTR7 repeat: matches 1..405 of consensus"
17661..18082
/note="MSMB repeat: matches 1..436 of consensus"
19420..19817
/note="MLT1B repeat: matches 6..390 of consensus"
21608..21667
/note="20 copies 3 mer gag 78% conserved"
22077..22227
/note="MLT1B repeat: matches 359..504 of consensus"
22323..22478
/note="78 copies 2 mer aa 59% conserved"
23255..24734
/note="L1MD repeat: matches -15..1314 of consensus"
24737..25041
/note="AluJo repeat: matches 1..296 of consensus"
25100..25570
/note="MLT1D repeat: matches 5..503 of consensus"
25807..25992
/note="AluSg/x repeat: matches 118..304 of consensus"
25993..26105
/note="MLT1-INTERNAL repeat: matches 1019..1121 of consensus"
26309..27011
/note="L1M4 repeat: matches 3454..4224 of consensus"
27012..27088
/note="5S repeat: matches 1..89 of consensus"
27094..27375
/note="L1M4 repeat: matches 4216..4507 of consensus"
27372..27878
/note="L1M1 repeat: matches 4771..5297 of consensus"
27891..28197
/note="AluJb repeat: matches 13..312 of consensus"
28219..28983
/note="MER21B repeat: matches 71..788 of consensus"
29017..29125
/note="L1MD1 repeat: matches 6116..6224 of consensus"
29805..29994
/note="MIR repeat: matches 59..251 of consensus"
30615..30873
/note="MIR repeat: matches 13..261 of consensus"
30916..30981
/note="22 copies 3 mer tgg 84% conserved"
31247..31537
/note="AluSx repeat: matches 23..301 of consensus"
31674..31970
/note="AluSc repeat: matches 1..298 of consensus"
32240..32562
/note="AluSx repeat: matches 1..312 of consensus"
33414..33457
/note="11 copies 4 mer acac 88% conserved"
34188..34484
/note="AluSx repeat: matches 1..303 of consensus"
34490..34614
/note="L1M1 repeat: matches 5433..5569 of consensus"
34616..34913
/note="AluSx repeat: matches 13..311 of consensus"

```


RESULT 3
AC098752/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-39G3, *** SEQUENCING IN PROGRESS ***, linear HTG 20-DEC-2001
AC098752 65064 bp DNA
AC098752 42 unordered pieces.
AC098752 2 GI:17972917
VERSION HTG; HTGS_PHASE1.
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE
AUTHORS 1 (bases 1 to 65064)
 Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C., Alsbrooks, S.L., Amaralunga, H.C., Are, J.R., Banks, T., Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulsegheh, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oah, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Reyes, M., Rojas, A., Rojokan, I., Rolfe, M., Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G. and Gibbs, R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 65064)
AUTHORS Worley, K.C.
TITLE Direct Submission
JOURNAL Submitted (01-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Dec 20, 2001 this sequence version replaced gi:16572781.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GGTL
 Center clone name: CH230-39G3
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329First call to flndPhrapList

Consensus quality: 51113 bases at least Q40
 Consensus quality: 54926 bases at least Q30
 Consensus quality: 57793 bases at least Q20
 Estimated insert size: 29709; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-gel estimation
 Quality coverage: 0.3x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently consists of 42 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1
 2591: contig of 2591 bp in length
 2691: gap of unknown length
 2692: contig of 2811 bp in length
 2692: gap of unknown length
 5503: contig of 1730 bp in length
 7332: gap of unknown length
 7332: gap of unknown length
 10138: contig of 2706 bp in length
 10139: gap of unknown length
 10238: contig of 1939 bp in length
 12177: gap of unknown length
 12178: contig of 2167 bp in length
 14444: gap of unknown length
 14445: contig of 1108 bp in length
 14545: gap of unknown length
 15652: contig of 2127 bp in length
 15752: gap of unknown length
 15753: contig of 1509 bp in length
 17880: gap of unknown length
 17880: contig of 1509 bp in length
 19488: gap of unknown length
 19489: contig of 2165 bp in length
 19589: gap of unknown length
 21754: contig of 1042 bp in length
 21853: gap of unknown length
 22895: contig of 1275 bp in length
 22895: gap of unknown length
 24270: contig of 1161 bp in length
 24271: gap of unknown length
 24370: contig of 1881 bp in length
 25531: gap of unknown length
 25531: contig of 1881 bp in length
 25632: gap of unknown length
 27512: contig of 1447 bp in length
 27512: gap of unknown length
 27613: contig of 1447 bp in length
 29059: gap of unknown length
 29059: contig of 1447 bp in length
 29060: gap of unknown length
 29159: contig of 1049 bp in length
 30208: gap of unknown length
 30208: contig of 1334 bp in length
 30309: gap of unknown length
 31642: contig of 1101 bp in length
 31742: gap of unknown length
 31743: contig of 1041 bp in length
 32843: gap of unknown length
 32844: contig of 1837 bp in length
 32944: gap of unknown length
 33984: gap of unknown length
 34084: contig of 1179 bp in length
 35921: gap of unknown length
 35921: contig of 1092 bp in length
 36021: gap of unknown length
 36022: contig of 1453 bp in length
 37001: gap of unknown length
 37001: contig of 1017 bp in length
 37201: gap of unknown length
 38392: contig of 1036 bp in length
 38392: gap of unknown length
 38492: contig of 1125 bp in length
 38945: gap of unknown length
 39945: gap of unknown length
 40045: contig of 1017 bp in length
 40045: gap of unknown length
 40046: contig of 1017 bp in length
 41062: gap of unknown length
 41062: contig of 1036 bp in length
 41162: gap of unknown length
 42298: contig of 1125 bp in length
 42299: gap of unknown length
 43424: gap of unknown length
 43424: contig of 1182 bp in length
 43524: gap of unknown length
 44706: contig of 1130 bp in length
 44805: gap of unknown length
 45935: gap of unknown length
 46035: gap of unknown length

* 46036 47316: contig of 1281 bp in length
 * 47317 47416: gap of unknown length
 * 47417 48893: contig of 1477 bp in length
 * 48894 48993: gap of unknown length
 * 48994 50069: contig of 1076 bp in length
 * 50070 50169: gap of unknown length
 * 50170 51313: contig of 1144 bp in length
 * 51314 54133: gap of unknown length
 * 54134 52416: contig of 1003 bp in length
 * 52417 52516: gap of unknown length
 * 52517 54086: contig of 1570 bp in length
 * 54087 54186: gap of unknown length
 * 54187 55690: contig of 1504 bp in length
 * 55691 55790: gap of unknown length
 * 55791 57121: contig of 1331 bp in length
 * 57122 57221: gap of unknown length
 * 57222 58288: contig of 1067 bp in length
 * 58289 58388: gap of unknown length
 * 58389 59001: contig of 1513 bp in length
 * 59002 60001: gap of unknown length
 * 60002 61061: contig of 1060 bp in length
 * 61062 61161: gap of unknown length
 * 61162 62494: contig of 1333 bp in length
 * 62495 62594: gap of unknown length
 * 62595 63732: contig of 1138 bp in length
 * 63733 63832: gap of unknown length
 * 63833 65064: contig of 1232 bp in length.

FEATURES

Location/Qualifiers
 1. 65064
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /clone="CH230-3963"

BASE COUNT 16376 a 14619 c 14087 g 15808 t 4174 others
 ORIGIN

Query Match 57.0%; Score 22.8; DB 2; Length 65064;

Best Local Similarity 79.4%; Pred. No. 50;

Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 ttcccaccatcatcccttggccctgggtgt 37

Db 21611 TTTCACCCCATATCCCTTTGGCTGCGCCTCT 21578

RESULT

AL355994/c
 LOCUS AL355994 157095 bp DNA linear HTG 10-AUG-2001
 DEFINITION Homo sapiens chromosome 1 clone RP11-254I4, *** SEQUENCING IN
 PROGRESS ***, 11 unordered pieces.

ACCESSION AL355994

VERSION AL355994.4 GI:10039855

KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.

SOURCE human

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 157095)

Mclay, K.

Direct Submission

Submitted (09-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Sep 9, 2000 this sequence version replaced gi:9588227.

----- Genome Center

Center: Sanger Centre

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: BA25414

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Consensus quality: 149592 bases at least Q40
 Consensus quality: 152642 bases at least Q30
 Consensus quality: 154560 bases at least Q20
 Insert size: 156095; sum-of-contigs
 Insert size: 156095; 3.9% error; agarose-fp
 Quality coverage: 4.04x in Q20 bases; sum-of-contigs Quality
 coverage: 4.13x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 11 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 2590: contig of 2590 bp in length
 * 2591 2690: gap of 100 bp
 * 2691 10620: contig of 7930 bp in length
 * 10621 10720: gap of 100 bp
 * 10721 39969: contig of 29249 bp in length
 * 39970 40689: gap of 100 bp
 * 40070 87137: contig of 47068 bp in length
 * 87138 87237: gap of 100 bp
 * 87238 99117: contig of 11880 bp in length
 * 99118 99217: gap of 100 bp
 * 99218 112196: contig of 12979 bp in length
 * 112197 112296: gap of 100 bp
 * 112297 116983: contig of 4687 bp in length
 * 116984 117083: gap of 100 bp
 * 117084 121151: contig of 4068 bp in length
 * 121152 121251: gap of 100 bp
 * 121252 134800: contig of 13549 bp in length
 * 134801 134900: gap of 100 bp
 * 134901 152651: contig of 17751 bp in length
 * 152652 152751: gap of 100 bp
 * 152752 157095: contig of 4344 bp in length.

FEATURES

source

1. 157095
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-254I4"
 /clone_lib="RPC1-11.1"

misc_feature

1. 2590
 /note="assembly_fragment:00613
 fragment_chain:1"

misc_feature

2691..10620
 /note="assembly_fragment:00394
 fragment_chain:1"

misc_feature

10721..39969
 /note="assembly_fragment:01099
 fragment_chain:1"

misc_feature

40070..87137
 /note="assembly_fragment:00143
 fragment_chain:1"

misc_feature

87238..99117
 /note="assembly_fragment:00340
 fragment_chain:2"

misc_feature

99218..112196
 /note="assembly_fragment:00372
 fragment_chain:2"

misc_feature

112297..116983
 /note="assembly_fragment:00457
 fragment_chain:3"

misc_feature

117084..121151
 /note="assembly_fragment:01157
 fragment_chain:3"

misc_feature

121252..134800
 /note="assembly_fragment:00625
 fragment_chain:4"

```

misc_feature 134901..152651
/note="assembly_fragment:01354
fragment_chain:4"
misc_feature 152752..157095
/note="assembly_fragment:01275
clone_end:SP6
vector_side:right"
BASE COUNT 37323 a 41621 c 41264 g 35885 t 1002 others
ORIGIN

Query Match 57.08; Score 22.8; DB 2; Length 157095;
Best Local Similarity 79.4%; Pred. No. 47;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 ctatttccaccatcccttggcccttggg 34
|||||
Db 76307 CTAATCCACCCAGCTCCCTGAGGCCCTGGG 76274

RESULT 5
MMU304861
LOCUS
DEFINITION MMU304861 24180 bp DNA linear ROD 21-DEC-2001
Mus musculus partial Bfsp2 gene for phakinin (CP49 protein), exon
1.
ACCESSION AJ304861
VERSION AJ304861.1 GI:17977855
KEYWORDS Bfsp2 gene; CP49 protein; intermediate filament protein; phakinin.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
AUTHORS Sandilands,A. and Quinlan,R.A.
TITLE Sequence analysis of the mouse CP49 gene
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 24180)
AUTHORS Sandilands,A.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2000) Sandilands A., Biochemistry, School of Life
Sciences, Dow Street, DD4 7AS, UNITED KINGDOM
FEATURES
source 1..24180
/organism="Mus musculus"
/strain="129"
/db_xref="taxon:10090"
7537..8028
/gene="Bfsp2"
7537..8028
/gene="Bfsp2"
/number=1
7537..8028
/gene="Bfsp2"
7537..8028
/gene="Bfsp2"
7537..8028
/gene="Bfsp2"
/function="intermediate filament protein"
/codon_start=1
/product="phakinin, CP49"
/protein_id="CAC83162.1"
/db_xref="GI:1797856"
/translation="MSKRRVAADLPSTNSMPVQRHVRVSSLRGTHSPSLDSPPPASR
TSAGSLVRAPGVYVAPSPGGGLGARVTRRALGISYVFLQGLRSSGLANYPAFCP
ERDHTTVDLGGCLVEYMTKVAHQVSOBLETLQRAHLESAKSGGMDALRASWAS
SYQQ"
BASE COUNT 6533 a 5929 c 5535 g 6183 t
ORIGIN

Query Match 56.5%; Score 22.6; DB 10; Length 24180;
Best Local Similarity 75.7%; Pred. No. 64;
Matches 28; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 ttccccaccatatacccttgggccccttgggtg 40
|||||
Db 4101 TGTCCCCCAAAACCCCTTGTTCCTGGTGGCGG 4137

RESULT 6
AC083820/c
LOCUS
DEFINITION AC083820 139572 bp DNA linear HTG 24-NOV-2001
Rattus norvegicus clone RP32-290H10, WORKING DRAFT SEQUENCE, 4
unordered pieces.
ACCESSION AC083820
VERSION AC083820.18 GI:17063037
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryote; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 139572)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Chen,R.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Focha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,D.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jollivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louleghed,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nkokenko,S.,
Ogih,M., Okuwonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N.,
Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
DIRECT SUBMISSION
Unpublished
REFERENCE 2 (bases 1 to 139572)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (01-OCT-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 24, 2001 this sequence version replaced gi:13346542.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: TUAIE

```

[illegible]

Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
 Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S.,
 Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.

TITLE JOURNAL REFERENCE AUTHORS JOURNAL

Direct Submission
 Unpublished
 2 (bases 1 to 67560)
 Worley, K.C.

TITLE

Submitted (17-Sep-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Dec 20, 2001 this sequence version replaced gi:15627285.

COMMENT

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GCYX
 Center clone name: CH230-8P8
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329first call to
 findPhrapList

Consensus quality: 54982 bases at least Q40
 Consensus quality: 59759 bases at least Q30
 Consensus quality: 63487 bases at least Q20
 Estimated insert size: 48793; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-fp estimation
 Quality coverage: 0.5x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 38 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 2811: contig of 2811 bp in length
 2812 2911: gap of unknown length
 2912 5574: contig of 2663 bp in length
 5575 5674: gap of unknown length
 5675 6898: contig of 1224 bp in length
 6899 6998: gap of unknown length
 6999 10205: contig of 3207 bp in length
 10206 10305: gap of unknown length
 10306 12288: contig of 1983 bp in length
 12289 12389: gap of unknown length
 12389 15169: contig of 2781 bp in length
 15170 15269: gap of unknown length
 15270 17310: contig of 2041 bp in length
 17311 17410: gap of unknown length
 17411 19518: contig of 2108 bp in length
 19519 19619: gap of unknown length
 19619 21450: contig of 1832 bp in length
 21451 21550: gap of unknown length
 21551 24016: contig of 2466 bp in length
 24017 24116: gap of unknown length
 24117 26516: contig of 2400 bp in length
 26517 26616: gap of unknown length
 26617 27921: contig of 1305 bp in length
 27922 28021: gap of unknown length
 28022 29551: contig of 1530 bp in length
 29552 29651: gap of unknown length
 29652 30966: contig of 1315 bp in length
 30967 31066: gap of unknown length
 31067 32926: contig of 1860 bp in length
 32927 33026: gap of unknown length
 33027 34991: contig of 1965 bp in length
 34992 35091: gap of unknown length

* 35092 36254: contig of 1163 bp in length
 36255 36354: gap of unknown length
 36355 37841: contig of 1487 bp in length
 37842 37941: gap of unknown length
 37942 39780: contig of 1839 bp in length
 39781 39880: gap of unknown length
 39881 40913: contig of 1033 bp in length
 40914 41013: gap of unknown length
 41014 42258: contig of 1245 bp in length
 42259 42358: gap of unknown length
 42359 43628: contig of 1270 bp in length
 43629 43729: gap of unknown length
 43730 45209: contig of 1481 bp in length
 45210 45309: gap of unknown length
 45310 46726: gap of unknown length
 46727 48161: contig of 1336 bp in length
 48162 48261: gap of unknown length
 48262 49952: contig of 1691 bp in length
 49953 50052: gap of unknown length
 50053 51149: contig of 1097 bp in length
 51150 51249: gap of unknown length
 51250 52985: contig of 1736 bp in length
 52986 53086: gap of unknown length
 53087 55192: contig of 2107 bp in length
 55193 55292: gap of unknown length
 55293 57279: contig of 1987 bp in length
 57280 57379: gap of unknown length
 57380 58527: contig of 1147 bp in length
 58528 58626: gap of unknown length
 58627 59848: contig of 1222 bp in length
 59849 61138: contig of 1190 bp in length
 61139 61239: gap of unknown length
 61240 62288: contig of 1049 bp in length
 62289 63440: contig of 1053 bp in length
 63441 63540: gap of unknown length
 63541 64590: contig of 1050 bp in length
 64591 64690: gap of unknown length
 64691 66334: contig of 1644 bp in length
 66335 66434: gap of unknown length
 66435 67560: contig of 1126 bp in length.

FEATURES source

1. 67560
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /clone="CH230-8P8"

BASE COUNT 19966 a 12041 c 11986 g 19717 t 3850 others
 ORIGIN

Query Match 56.0%; Score 22.4; DB 2; Length 67560;
 Best Local Similarity 72.5%; Pred. No. 73;
 Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Oy 1 ctatttccaccatattcccttgcccttggtgggtggg 40
 ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 56938 CTCCTTCCCAACCCATTCCTTGTGTGAGAAGG 56899

RESULT 10 AC102525/c

LOCUS AC102525 70165 bp DNA linear HTG 23-NOV-2001
 DEFINITION MUS musculus clone RP24-133K20, LOW-PASS SEQUENCE SAMPLING.
 ACCESSION AC102525
 VERSION AC102525.1 GI:17061611
 KEYWORDS HTG; HTGS_PHASE0.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 REFERENCE 1 (bases 1 to 70165)

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Mus musculus, clone RP24-133K20
2 (bases 1 to 70165)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choepe, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A.,
Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Hearford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., Lakocque, K.,
Lamacz, R., Landers, T., Lehotzky, J., Levine, R., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J.,
Meneau, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L18959
Center clone name: 133_K20

* NOTE: This record contains 89 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 671: contig of 671 bp in length
* 672 771: gap of 100 bp
* 772 1459: contig of 688 bp in length
* 1460 1559: gap of 100 bp
* 1560 2254: contig of 695 bp in length
* 2255 2354: gap of 100 bp
* 2355 3037: contig of 673 bp in length
* 3038 3127: gap of 100 bp
* 3128 3834: contig of 707 bp in length
* 3835 3934: gap of 100 bp
* 3935 4634: contig of 700 bp in length
* 4635 4734: gap of 100 bp
* 4735 5433: contig of 699 bp in length
* 5434 5533: gap of 100 bp
* 5534 6232: contig of 699 bp in length
* 6233 6332: gap of 100 bp
* 6333 7034: contig of 702 bp in length
* 7035 7134: gap of 100 bp
* 7135 7842: contig of 708 bp in length
* 7843 7942: gap of 100 bp
* 7943 8604: contig of 662 bp in length
* 8605 8704: gap of 100 bp
* 8705 9399: contig of 695 bp in length
* 9400 9499: gap of 100 bp
* 9500 10193: gap of 100 bp
* 10194 10993: gap of 100 bp
* 10994 10986: contig of 693 bp in length
* 10987 11086: gap of 100 bp
* 11087 11759: contig of 673 bp in length
* 11760 11859: gap of 100 bp
* 11860 12555: contig of 696 bp in length
* 12556 12555: gap of 100 bp
* 12556 13343: contig of 688 bp in length
* 13344 13443: gap of 100 bp
* 13444 14147: contig of 704 bp in length
* 14148 14247: gap of 100 bp
* 14248 14930: contig of 683 bp in length
* 14931 15030: gap of 100 bp
* 15031 15702: contig of 672 bp in length
* 15703 15802: gap of 100 bp
* 15803 16477: contig of 675 bp in length
* 16378 16577: gap of 100 bp
* 16378 17253: contig of 676 bp in length
* 17254 17353: gap of 100 bp
* 17354 18045: contig of 692 bp in length
* 18046 18145: gap of 100 bp
* 18146 18840: contig of 695 bp in length
* 18941 18940: gap of 100 bp
* 18941 19616: contig of 676 bp in length
* 19617 19716: gap of 100 bp
* 19717 20423: contig of 707 bp in length
* 20424 20523: gap of 100 bp
* 20524 21200: contig of 677 bp in length
* 21201 21300: gap of 100 bp
* 21301 22001: contig of 701 bp in length
* 22002 22101: gap of 100 bp
* 22102 22799: contig of 698 bp in length
* 22800 22999: gap of 100 bp
* 22900 23597: contig of 698 bp in length
* 23598 23697: gap of 100 bp
* 23698 24396: contig of 699 bp in length
* 24397 24496: gap of 100 bp
* 24497 25167: contig of 671 bp in length
* 25168 25267: gap of 100 bp
* 25268 25956: contig of 689 bp in length
* 25957 26056: gap of 100 bp
* 26057 26742: contig of 686 bp in length
* 26743 26842: gap of 100 bp
* 26843 27534: contig of 692 bp in length
* 27535 27634: gap of 100 bp
* 27635 28323: contig of 689 bp in length
* 28324 28423: gap of 100 bp
* 28424 29134: contig of 711 bp in length
* 29135 29234: gap of 100 bp
* 29235 29882: contig of 648 bp in length
* 29883 29982: gap of 100 bp
* 29983 30672: contig of 690 bp in length
* 30673 30772: gap of 100 bp
* 30773 31476: contig of 704 bp in length
* 31477 31576: gap of 100 bp
* 31577 32258: contig of 682 bp in length
* 32259 32358: gap of 100 bp
* 32359 33040: contig of 682 bp in length
* 33041 33140: gap of 100 bp
* 33141 33816: contig of 676 bp in length
* 33817 33916: gap of 100 bp
* 33917 34584: contig of 668 bp in length
* 34585 34684: gap of 100 bp
* 34685 35360: contig of 676 bp in length
* 35361 35460: gap of 100 bp
* 35461 36136: contig of 676 bp in length
* 36137 36236: gap of 100 bp
* 36237 36932: contig of 696 bp in length
* 36933 37032: gap of 100 bp

* 37033 37713: contig of 681 bp in length
* 37714 37813: gap of 100 bp
* 37814 38521: contig of 708 bp in length
* 38522 38621: gap of 100 bp
* 38622 39323: contig of 702 bp in length
* 39324 39423: gap of 100 bp
* 39424 40125: contig of 702 bp in length
* 40126 40225: gap of 100 bp
* 40226 40927: contig of 702 bp in length
* 40928 41027: gap of 100 bp
* 41028 41723: contig of 696 bp in length
* 41724 41823: gap of 100 bp
* 41824 42525: contig of 702 bp in length
* 42526 42625: gap of 100 bp
* 42626 43300: contig of 675 bp in length
* 43301 43400: gap of 100 bp
* 43401 44101: contig of 701 bp in length
* 44102 44201: gap of 100 bp
* 44202 44901: contig of 700 bp in length
* 44902 45001: gap of 100 bp
* 45002 45691: contig of 690 bp in length
* 45692 45791: gap of 100 bp
* 45792 46494: contig of 703 bp in length
* 46495 46594: gap of 100 bp
* 46595 47289: contig of 695 bp in length
* 47290 47389: gap of 100 bp
* 47390 48066: contig of 677 bp in length
* 48067 48166: gap of 100 bp
* 48167 48864: contig of 698 bp in length
* 48865 48964: gap of 100 bp
* 48965 49663: contig of 699 bp in length
* 49664 49763: gap of 100 bp
* 49764 50442: contig of 679 bp in length
* 50443 50542: gap of 100 bp
* 50543 51235: contig of 693 bp in length
* 51236 51335: gap of 100 bp
* 51336 52022: contig of 687 bp in length
* 52023 52122: gap of 100 bp
* 52123 52826: contig of 704 bp in length
* 52827 52926: gap of 100 bp
* 52927 53608: contig of 682 bp in length
* 53609 53708: gap of 100 bp
* 53709 54394: contig of 686 bp in length
* 54395 54494: gap of 100 bp

Query Match 56.0%; Score 22.4; DB 2; Length 70165;
Best Local Similarity 78.8%; Pred. No. 73;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 tattttcccccacatccccctgggccccttggg 34
| | | | | | | | | | | | | | | | | | | | | |
Db 29060 TTTTCCACCCACATCCCCCTGCTCTGGG 29028

RESULT 11
AL136523
LOCUS Human DNA sequence from clone RP11-103K9 on chromosome 13, complete
DEFINITION sequence.
ACCESSION AL136523.20 GI:15384799
VERSION HTG.
KEYWORDS Homo sapiens
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 86859)
AUTHORS Tromans,A.
TITLE Direct Submission
JOURNAL Submitted (23-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerquest@sanger.ac.uk
COMMENT On Aug 31, 2001 this sequence version replaced gi:15021280.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGF/Chr13
RP11-103K9 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6
IMPORTANT: This sequence is not the entire insert of clone RP11-103K9 It may be shorter because we sequence overlapping sections only once, except for a short overlap.
The true left end of clone RP11-431022 is at 84860 in this sequence. The true right end of clone RP11-209J19 is at 2000 in this sequence.

FEATURES
source
1. 86859
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-103K9"
/clone_lib="RPCI-11.1"
complement(261..577)
/note="match: STS: Em:HSPH11F3"
complement(19452..19963)
/note="match: STS: Em:G63496"
45385..45406
/note="Sequence from uni-directional dGTP big dye terminator reads only."
BASE COUNT 24999 a 20505 c 18043 g 23312 t
ORIGIN

Query Match 56.0%; Score 22.4; DB 9; Length 86859;
Best Local Similarity 72.5%; Pred. No. 72;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 ctatttcccccacatccccctgggccccttgggtgtggg 40
| | | | | | | | | | | | | | | | | | | | | |
Db 62785 CTATTTCACACTATCCCGCTGGCCCTAGATTGTGGG 62824

RESULT 12
AP000005/c
LOCUS Homo sapiens genomic DNA, chromosome 21q22.1, segment 27/28,
DEFINITION complete sequence.
ACCESSION AP0000056
VERSION AP0000056.1 GI:3132366
KEYWORDS HTG.
SOURCE Homo sapiens DNA, clone:245w17-f4A4f_6.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 100000)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Homo sapiens genomic DNA, chromosome 21q
 2 (bases 1 to 100000)
 Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
 Direct Submission
 Submitted (11-MAY-1998) to the DDBJ/EMBL/GenBank databases.
 Masahira Hattori, Kitasato University, Department of Science, JST
 Sequencing Laboratory, Kitasato 1-15-1, Sagamihara 228, Japan
 (E-mail:hattori@hgc.ims.u-tokyo.ac.jp, Tel:0427-78-9732,
 Fax:0427-78-9561)
 This sequence is conducted by Kitasato University JST sequencing
 Laboratory as a JST sequencing team.
 Principal Investigator:Yoshiyuki Sakaki Ph.D.
 Phone: +81-3-5449-5622, Fax : +81-3-5449-5445,
 sakaki@hgc.ims.u-tokyo.ac.jp
 Sub-leader: Tadayoshi Shiba Ph.D., Masahira Hattori Ph.D. The
 sequence is submitted by: Human Genome Sequencing in ALIS project of
 JST

COMMENT

Japan Science and Technology Corporation (JST)
 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0028 Japan
 For further information about this sequence, including its location
 and relationship to other sequences, please visit our sequence
 archive Web site (<http://www-alis.tokyo.jst.go.jp/HGS/top.html>)
 or send email to webmaster@www-alis.tokyo.jst.go.jp.

FEATURES

source
 1. .100000
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="21"
 /clone="245P17-f4A4f_6"
 /map="21q22.1"
 /map="21q22.1"

BASE COUNT 26660 a 22366 c 22356 g 28618 t

ORIGIN

Query Match 56.08; Score 22.4; DB 9; Length 100000;
Best Local Similarity 72.5%; Pred. No. 71;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 ctatttccaccatatacccttggcccttgggtggtggg 40
 ||||| || ||||| || ||||| || ||||| || |||||

Db 71243 CCATTGCTTCGCATATACAGTTGGCCCTTGAATGTTGG 71204

RESULT 13

AP000124/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AP000124 100000 bp DNA linear PRI 25-SEP-1999
 Homo sapiens genomic DNA of 21q22.1, GART and AML related,
 SLC5A3-f4A4 region, segment 7/8, complete sequence.
 AP000124
 AP000124.1 GI:4730893
 HTG.
 Homo sapiens DNA.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheraia; Primates; Catarrhini; Hominidae; Homo.
 1 (sites)
 Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
 Homo sapiens 817,199bp genomic DNA of 21q22.1 GART and AML region
 Published Only in Database (1999) In press
 2 (bases 1 to 100000)
 Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,J.
 Direct Submission
 Submitted (15-APR-1999) to the DDBJ/EMBL/GenBank databases. Miki
 Hirakawa, Japan Science and Technology Corporation (JST), Advanced
 Databases Department, 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081,
 Japan (E-mail:mika@tokyo.jst.go.jp,
 URL:<http://www-alis.tokyo.jst.go.jp>, Tel:81-3-5214-8491,
 Fax:81-3-5214-8470)
 This sequence is conducted by Kitasato University JST sequencing
 Laboratory as a JST sequencing team.
 Principal Investigator:Yoshiyuki Sakaki Ph.D.
 Phone: +81-3-5449-5622, Fax : +81-3-5449-5445,

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Submitted (10-MAY-1999) to the DDBJ/EMBL/GenBank databases.
 Masahira Hattori, The Institute of Physical and Chemical Research
 (RIKEN), Genomic Sciences Center (GSC), Kitasato Univ., 1-15-1
 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
 (E-mail:hattori@hgc.riken.go.jp, URL:<http://hgc.gsc.riken.go.jp/>,
 Tel:81-42-778-9923, Fax:81-42-778-9924)
 E. coli transposon insertion: The present data does not contain E.
 coli transposon sequences which integrated in the
 original/previous sequences. We determined the boundary between
 the insertion and genomic sequences experimentally, removed the
 insertion sequences, reconstituted the present data. The sequencing
 project is supported by Japan Science Technology Corporation (JST)
 and The Institute of Physical and Chemical Research (RIKEN).
 Location/Qualifiers
 1. .100000
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

sakaki@hgc.ims.u-tokyo.ac.jp
 Sub-leader: Tadayoshi Shiba Ph.D., Masahira Hattori Ph.D. The
 sequence is submitted by Human Genome Sequencing in ALIS project of
 JST.
 Japan Science and Technology Corporation (JST)
 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081 Japan
 For further information about this sequence, including its location
 and relationship to other sequences, please visit our sequence
 archive Web site (<http://www-alis.tokyo.jst.go.jp/HGS/>) or send
 email to webmaster@www-alis.tokyo.jst.go.jp.

FEATURES

source
 1. .100000
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="21"
 /map="21q22.1"
 /map="21q22.1"
 /standard_name="D21S65"
 /note="D21S65.PCR1; The location is between each flanking
 site of PCR primers."
 /db_xref="GDB:191986"
 /db_xref="GDB:191986"

BASE COUNT 26604 a 22172 c 22427 g 28797 t

ORIGIN

Query Match 56.08; Score 22.4; DB 9; Length 100000;
Best Local Similarity 72.5%; Pred. No. 71;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 ctatttccaccatatacccttggcccttgggtggtggg 40
 ||||| || ||||| || ||||| || ||||| || |||||

Db 80545 CCATTGCTTCGCATATACAGTTGGCCCTTGAATGTTGG 80506

RESULT 14

AP000171/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AP000171 100000 bp DNA linear PRI 08-JAN-2000
 Homo sapiens genomic DNA, chromosome 21q22.1, D21S226-AML region,
 clone B2344F14-f50E8, segment 7/9, complete sequence.
 AP000171
 AP000171.1 GI:4827136
 HTG.
 Homo sapiens DNA.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheraia; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 100000)
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.
 Homo sapiens 890,291bp genomic DNA of 21q22.1 (REGION: D21S226-AML
 CLONE RANGE: B2344F14-f50E8)
 Published Only in Database (1999) In press
 2 (bases 1 to 100000)
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.
 Direct Submission
 Submitted (10-MAY-1999) to the DDBJ/EMBL/GenBank databases.
 Masahira Hattori, The Institute of Physical and Chemical Research
 (RIKEN), Genomic Sciences Center (GSC), Kitasato Univ., 1-15-1
 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
 (E-mail:hattori@hgc.riken.go.jp, URL:<http://hgc.gsc.riken.go.jp/>,
 Tel:81-42-778-9923, Fax:81-42-778-9924)
 E. coli transposon insertion: The present data does not contain E.
 coli transposon sequences which integrated in the
 original/previous sequences. We determined the boundary between
 the insertion and genomic sequences experimentally, removed the
 insertion sequences, reconstituted the present data. The sequencing
 project is supported by Japan Science Technology Corporation (JST)
 and The Institute of Physical and Chemical Research (RIKEN).
 Location/Qualifiers
 1. .100000
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Submitted (10-MAY-1999) to the DDBJ/EMBL/GenBank databases.
 Masahira Hattori, The Institute of Physical and Chemical Research
 (RIKEN), Genomic Sciences Center (GSC), Kitasato Univ., 1-15-1
 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
 (E-mail:hattori@hgc.riken.go.jp, URL:<http://hgc.gsc.riken.go.jp/>,
 Tel:81-42-778-9923, Fax:81-42-778-9924)
 E. coli transposon insertion: The present data does not contain E.
 coli transposon sequences which integrated in the
 original/previous sequences. We determined the boundary between
 the insertion and genomic sequences experimentally, removed the
 insertion sequences, reconstituted the present data. The sequencing
 project is supported by Japan Science Technology Corporation (JST)
 and The Institute of Physical and Chemical Research (RIKEN).
 Location/Qualifiers
 1. .100000
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

COMMENT

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 17:08:55 ; Search time 4272.65 Seconds
(without alignments)
126.357 Million cell updates/sec

Title: US-09-721-543A-14
Perfect score: 40
Sequence: 1 catacctgaccctactctacctggcgtgactgggtggg 40

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: em_estha:*
2: em_esthum:*
3: em_estin:*
4: em_estinu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	24.2	60.5	534	12	TA297H02Q
2	23.8	59.5	217	10	BF987540
3	23.8	59.5	687	12	AG112675
c	22.2	55.5	334	10	BF813535
5	22.2	55.5	390	12	A2050757
6	22.2	55.0	520	12	A2856792
7	21.8	54.5	559	10	B1538782
8	21.8	54.5	636	10	B1659337
9	21.6	54.0	230	9	A1629579
10	21.6	54.0	343	9	A1586386
11	21.6	54.0	350	9	A1920579
12	21.6	54.0	421	9	A1438418
13	21.6	54.0	422	9	A1600794
14	21.6	54.0	426	9	A1065660
c	15	21.6	54.0	437	AW917931
16	21.6	54.0	462	10	B1388746
17	21.6	54.0	474	9	AW091238

18	21.6	54.0	486	9	AI987339	AI987339	660004B08
19	21.6	54.0	493	9	AI065928	aj44h01.x	AI065928
20	21.6	54.0	501	9	AW360562	660035F05	AW360562
21	21.6	54.0	510	9	AI665059	605006C07	AI665059
22	21.6	54.0	544	9	AI691732	606007F07	AI691732
23	21.6	54.0	558	10	B1542879	949074A05	B1542879
24	21.6	54.0	565	12	TA92D07Q	T. brucei	AI462429
25	21.6	54.0	567	9	AI622635	486105A04	AI622635
26	21.6	54.0	569	9	AI739761	606050E03	AI739761
27	21.6	54.0	573	9	AI795446	605006C07	AI795446
28	21.6	54.0	577	9	AI714407	606013D02	AI714407
29	21.6	54.0	578	9	AW225318	687056F02	AW225318
30	21.6	54.0	581	10	BM335923	MEST179-A	BM335923
31	21.6	54.0	583	9	AI734339	606029G11	AI734339
32	21.6	54.0	595	9	AW331631	687044C07	AW331631
33	21.6	54.0	596	10	BM078367	MEST118-G	BM078367
34	21.6	54.0	599	9	AI670159	605019B06	AI670159
35	21.6	54.0	617	9	AI677241	605050C08	AI677241
36	21.6	54.0	679	9	AI676891	605046D05	AI676891
37	21.4	53.5	298	9	AI154465	ua02h01.x	AI154465
38	21.4	53.5	301	10	BF896828	PM3-MT011	BF896828
39	21.4	53.5	418	10	W66884	me20b09.r1	W66884
40	21.4	53.5	431	12	A0722584	HS_5236.A	A0722584
41	21.4	53.5	443	9	AA125134	mp78C05.F	AA125134
42	21.4	53.5	466	9	AI180661	uc47c10.r	AI180661
43	21.4	53.5	488	10	BM232888	R0329E04-	BM232888
44	21.4	53.5	491	10	BE458863	CV330808.Y	BE458863
45	21.4	53.5	502	10	BF986408	OVA-GN014	BF986408

ALIGNMENTS

RESULT	1
TA297H02Q	534 bp DNA linear GSS 13-DEC-2000
LOCUS	T. brucei sheared genomic DNA clone 297h02, reverse sequence, genomic survey sequence.
DEFINITION	genomic survey sequence.
ACCESSION	AL489616.1 GI:11864911
VERSION	GSS
KEYWORDS	Trypanosoma brucei.
SOURCE	Trypanosoma brucei
ORGANISM	Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae;
REFERENCE	1 (bases 1 to 534)
AUTHORS	Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R., Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L., Melville, S.E., Rajandream, M.A. and Barrell, B.G.
TITLE	Direct Submission
JOURNAL	Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nhlsanger.ac.uk
COMMENT	Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999). Email: nelsayed@tigr.org
FEATURES	Details of T. brucei sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/projects/T-brucei/ .
Source	Location/Qualifiers
	1..534
	/organism="Trypanosoma brucei"
	/strain="TREU927"
	/db_xref="taxon:5691"
	/clone="297H02"
BASE COUNT	135 a 117 c 135 g 147 t

ORIGIN

Query Match 60.5%; Score 24.2; DB 12; Length 534;
 Best Local Similarity 78.4%; Pred. No. 42;
 Matches 29; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 2 atcacttgaccctactctacctggctggactgggtg 38
 ||||| |||| ||||| ||||| ||||| ||||| |||||
 Db 439 ATCACTTTACCCACCTCTACTTGGGCTGAGCTGCGG 475

RESULT 2

BF987540 217 bp mRNA linear EST 23-JAN-2001
 DEFINITION MR3-GN0150-101000-001-b12 GN0150 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BF987540
 VERSION BF987540.1 GI:12339862
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 217)

REFERENCE Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
 Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663

COMMENT Contact: Simpson A.J.J.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR3st2=MR3-GN0150-101000-001-b12&t3=2000-10-10&t4=1>)
 Seq primer: puc 18 forward
 High quality sequence start: 24
 High quality sequence stop: 217.

FEATURES

source
 1..217
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="GN0150"
 /dev_stage="Adult"

/note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI
 ; Site_2: SmaI; A mini-library was made by cloning
 products derived from ORESTES PCR (U.S. Letters Patent
 application No. 196,716 - Ludwig Institute for Cancer
 Research) profiles into the pUC 18 vector. Reverse
 transcription of tissue mRNA and cDNA amplification were
 performed under low stringency conditions."

BASE COUNT 47 a 44 c 81 g 45 t

ORIGIN

Query Match 59.5%; Score 23.8; DB 10; Length 217;
 Best Local Similarity 80.0%; Pred. No. 46;
 Matches 28; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 3 tcacttgaccctactctacctggctggactgggtg 37
 ||||| |||| ||||| ||||| ||||| ||||| |||||
 Db 170 TCACGTGACCAATCTCTAGCTGGGCTGGACAGGTT 204

RESULT 3

AG112675 687 bp DNA linear GSS 03-NOV-2001
 LOCUS Pan troglodytes DNA, clone: PTB-119806.F, genomic survey sequence.
 ACCESSION AG112675
 VERSION AG112675.1 GI:16733194
 KEYWORDS GSS; GSS (genome survey sequence).
 SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
 BAC Library clone:PTB-119806.F.

ORGANISM

Pan troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
 1 (sites)
 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
 Totoki, Y., Watanabe, H. and Sakaki, Y.
 BAC end sequences of Library PTB

REFERENCE

2 (bases 1 to 687)
 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
 Totoki, Y., Watanabe, H. and Sakaki, Y.
 Direct Submission

TITLE

JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 1-7-22 Suehiro-chou, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail: chimbres@gsc.riken.go.jp, URL: <http://hgp.gsc.riken.go.jp/>,
 Tel: 81-45-503-9111 Fax: 81-45-503-9170)

Clones are derived from the chimpanzee BAC library PTB This BAC end
 was generated during the R&D process and may have higher chance of
 clone tracking errors.

PRIMERS

Sequencing: -21M13

LIBRARY

Vector : pKS145
 R.Site 1 : SacI
 R.Site 2 : SacI.
 Location/Qualifiers
 1..687
 /organism="Pan troglodytes"
 /db_xref="taxon:9598"
 /clone="PTB-119806.F"
 /sex="male"
 /cell_type="lymphoblast"
 /clone_lib="PTB Chimpanzee Male BAC Library"
 192 a 132 c 234 g 128 t 1 others

FEATURES

source

1..687

/organism="Pan troglodytes"

/db_xref="taxon:9598"

/clone="PTB-119806.F"

/sex="male"

/cell_type="lymphoblast"

/clone_lib="PTB Chimpanzee Male BAC Library"

192 a 132 c 234 g 128 t 1 others

BASE COUNT 192 a 132 c 234 g 128 t 1 others

ORIGIN

Query Match 59.5%; Score 23.8; DB 12; Length 687;

Best Local Similarity 80.0%; Pred. No. 64;

Matches 28; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 3 tcacttgaccctactctacctggctggactgggtg 37

||||| |||| ||||| ||||| ||||| ||||| |||||

Db 82 TCACGTGACCAATCTCTAGCTGGGCTGGACAGGTT 116

||||| |||| ||||| ||||| ||||| ||||| |||||

RESULT 4

BF813535/c

LOCUS BF813535

DEFINITION ILS-C10148-271100-277-e05 C10148 Homo sapiens cDNA, mRNA sequence.

ACCESSION BF813535

VERSION BF813535.1 GI:12144616

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 334)

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,

Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,

Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,

Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,

Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,

Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,

Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,

Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,

Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,

Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,

Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,

Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,

JOURNAL COMMENT
 Unpublished (1999)
 Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plates: 486048 row: B column: 02.
 Location/Qualifiers
 1..343
 /organism="Zea mays"
 /cultivar="B73"
 /db_xref="taxon:4577"
 /clone_lib="486 - leaf primordia cDNA library from Hake lab"
 /tissue_type="leaf primordia"
 /dev_stage="P7-P11 leaf"
 /lab_host="E.coli XL1-Blue MFR"
 /note="Organ: shoot; Vector: Lambda zap; Hake lab cDNA library."
 BASE COUNT 93 a 73 c 63 g 114 t
 ORIGIN
 Query Match 54.0%; Score: 21.6; DB 9; Length 343;
 Best Local Similarity 75.0%; Pred. No: 3.6e+02;
 Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
 Qy 5 acctgaccctactctacctggctggactggctgggtggg 40
 ||||| | | ||||| | | | | | | | | | | | | | | | |
 Db 257 ACTTGACCAATCCTGACTGGCTGGACCGGGAGCG 292
 RESULT 11
 AI920579
 LOCUS 618015G10.x1 618 - Inbred Tassel cDNA Library Zea mays cDNA, mRNA
 DEFINITION sequence.
 ACCESSION AI920579.1 GI:5650219
 VERSION AI920579
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Zea.
 REFERENCE 1 (bases 1 to 350)
 AUTHORS Walbot,V.
 TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
 JOURNAL University
 COMMENT Unpublished (1999)
 Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 618015 row: G column: 10.
 Location/Qualifiers
 1..350
 /organism="Zea mays"
 /cultivar="Ohio43"
 /db_xref="taxon:4577"
 /clone_lib="618 - Inbred Tassel cDNA Library"
 /tissue_type="tassel"
 /dev_stage="tassel length from 0.1 to 2.5 cm"
 /lab_host="XLOLR"
 /note="Organ: tassel; Vector: pAD-GAL4-2.1 (Hybrizap);
 Inbred tassel library from Schmidt lab"
 BASE COUNT 101 a 82 c 80 g 87 t
 ORIGIN

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 17:47:11 : Search time 116.71 Seconds
(without alignments)
84.186 Million cell updates/sec

Title: US-09-721-543A-14
Perfect score: 40
Sequence: 1 catacattgaccctactctacctggctgactgggtggg 40

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/lna/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/lna/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/lna/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/lna/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/lna/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/lna/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	20.2	50.5	359	4	US-09-085-199B-30
C 2	20.2	50.5	1164	4	US-09-085-199B-1
C 3	20.2	50.5	4796	4	US-09-085-199B-3
4	19.4	48.5	56516	2	US-08-996-306-1
5	19.4	48.5	56516	4	US-09-338-907-1
6	19.4	48.5	56516	4	US-09-218-207-1
7	19.4	48.5	56520	4	US-09-338-907-179
8	19.4	48.5	56520	4	US-09-218-207-179
9	19.2	48.0	2389	1	US-08-123-161A-13
10	19.2	48.0	2389	1	US-08-483-278-13
11	19	47.5	969	3	US-09-188-930-11
12	19	47.5	2515	1	US-08-061-465-3
13	19	47.5	4002	2	US-08-231-193A-53
14	19	47.5	4002	2	US-08-486-273A-53
15	19	47.5	4002	3	US-08-480-474-53
16	19	47.5	4002	3	US-08-940-086A-53
17	19	47.5	4002	4	US-08-940-035A-53
18	19	47.5	4017	2	US-08-231-193A-49
19	19	47.5	4017	2	US-08-486-273A-49
20	19	47.5	4017	3	US-08-480-474-49
21	19	47.5	4017	3	US-08-940-086A-49
22	19	47.5	4017	3	US-08-940-035A-49
23	19	47.5	4053	2	US-08-231-193A-47
24	19	47.5	4053	2	US-08-486-273A-47
25	19	47.5	4053	3	US-08-480-474-47
26	19	47.5	4053	3	US-08-940-086A-47
27	19	47.5	4053	4	US-08-940-035A-47

28	19	47.5	4068	2	US-08-231-193A-5	Sequence 5, Appli
29	19	47.5	4068	2	US-08-486-273A-5	Sequence 5, Appli
30	19	47.5	4068	3	US-08-480-474-5	Sequence 5, Appli
31	19	47.5	4068	3	US-08-940-086A-5	Sequence 5, Appli
32	19	47.5	4068	4	US-08-940-035A-5	Sequence 5, Appli
33	19	47.5	4077	2	US-08-231-193A-51	Sequence 51, Appl
34	19	47.5	4077	2	US-08-486-273A-51	Sequence 51, Appl
35	19	47.5	4077	3	US-08-480-474-51	Sequence 51, Appl
36	19	47.5	4077	3	US-08-940-086A-51	Sequence 51, Appl
37	19	47.5	4077	4	US-08-940-035A-51	Sequence 51, Appl
38	19	47.5	4092	2	US-08-231-193A-45	Sequence 45, Appl
39	19	47.5	4092	2	US-08-486-273A-45	Sequence 45, Appl
40	19	47.5	4092	3	US-08-480-474-45	Sequence 45, Appl
41	19	47.5	4092	3	US-08-940-086A-45	Sequence 45, Appl
42	19	47.5	6828	1	US-08-061-465-1	Sequence 1, Appli
43	19	47.5	6828	1	US-08-940-035A-45	Sequence 11, Appl
44	18.8	47.0	1149	4	US-09-434-774-11	Sequence 1, Appli
45	18.8	47.0	72928	3	US-09-009-913-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-085-199B-30/c
; Sequence 30, Application US/09085199B
; Patent No. 6235879
; GENERAL INFORMATION:
; APPLICANT: Hayden, Michael R.
; APPLICANT: Hackam, Abigail
; APPLICANT: Huq, A.H.M. Mahbubul
; APPLICANT: Chopra, Vikramjit Singh
; APPLICANT: Kalchman, Michael
; TITLE OF INVENTION: Apoptosis Modulators That Interact with the
; TITLE OF INVENTION: Huntington's Disease Gene
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oppedahl & Larson
; STREET: PO Box 5270
; CITY: Frisco
; STATE: CO
; COUNTRY: USA
; ZIP: 80443-5270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Kb storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS DOS 5.0
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/085,199B
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Larson, Marina T.
; REGISTRATION NUMBER: 32038
; REFERENCE/DOCKET NUMBER: UBC.P-013US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (970) 668-2050
; TELEFAX: (970) 668-2052
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 359
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; HYPOTHETICAL: no
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; ORGANISM: human
; FEATURE:
; OTHER INFORMATION: exon 15 of HTP1

US-09-085-199B-30

Query Match 50.5%; Score 20.2; DB 4; Length 359;
Best Local Similarity 75.8%; Pred. No. 14;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 3 tcacttgaccctactactacctggcgtgactgg 35
 ||| | | | | | | | | |
Db 188 TCTCTCGTTCCAAATACCTGGGCTGTGTGG 156

RESULT 2

```

US-09-085-199B-1/C
; Sequence 1, Application US/09085199B
; Patent No. 6235879
; GENERAL INFORMATION:
; APPLICANT: Hayden, Michael R.
; APPLICANT: Hackam, Abigail
; APPLICANT: Huq, A.H.M. Mahbubul
; APPLICANT: Chopra, Vikramjit Singh
; APPLICANT: Kachman, Michael
; TITLE OF INVENTION: Apoptosis Modulators That Interact with the
; TITLE OF INVENTION: Huntington's Disease Gene
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oppedahl & Larson
; STREET: PO Box 5270
; CITY: Frisco
; STATE: CO
; COUNTRY: USA
; ZIP: 80443-5270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 Inch, 1.44 Kb storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS DOS 5.0
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/085,199B
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Larson, Marina T.
; REGISTRATION NUMBER: 32038
; REFERENCE/DOCKET NUMBER: UBC.P-013US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (970) 668-2050
; TELEFAX: (970) 668-2052
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1164
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: no
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; ORGANISM: human
; FEATURE: cDNA for Huntingtin-interacting protein
US-09-085-199B-1

```

RESULT 3

RESUL 3
US-09-085-199B-3/c

```

: Sequence 3, Application US/09085199B
: Patent No. 6235879
: GENERAL INFORMATION:
: APPLICANT: Hayden, Michael R.
: APPLICANT: Hackam, Abigail
: APPLICANT: Huq, A.H.M. Mahbubul
: APPLICANT: Chopra, Vikramjit Singh
: APPLICANT: Kalchman, Michael
: TITLE OF INVENTION: Apoptosis Modulators That Interact with the
: TITLE OF INVENTION: Huntington's Disease Gene
: NUMBER OF SEQUENCES: 44
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Oppedani & Larson
: STREET: PO Box 5270
: CITY: Frisco
: STATE: CO
: COUNTRY: USA
: ZIP: 80443-5270
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Kb storage
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: MS DOS 5.0
: SOFTWARE: Wordperfect
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/085.199B
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Larson, Marina T.
: REGISTRATION NUMBER: 32038
: REFERENCE/DOCKET NUMBER: UBC.P-013US2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (970) 668-2050
: TELEFAX: (970) 668-2052
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4796
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHETICAL: no
: ANTI-SENSE: no
: ORIGINAL SOURCE:
: ORGANISM: human
: FEATURE:
: OTHER INFORMATION: cDNA for Huntingon-interacting protein
: US-09-085-199B-3

```

RESULT

```

4
US-08-996-306-1
: Sequence 1, Application US/08996306
: Patent No. 5945522
: GENERAL INFORMATION:
: APPLICANT: Cohen, Daniel
: APPLICANT: Chumakov, Ilya
: APPLICANT: Blumenfeld, Marta
: APPLICANT: Bougueleret, Lydie
: TITLE OF INVENTION: Prostate cancer gene
: NUMBER OF SEQUENCES: 68
: CORRESPONDENCE ADDRESS:
: ADDRESS: Knobbe, Martens, Olson & Bear
: STREET: 501 West Broadway
:

```


SOFTWARE: Patent.pm
; SEQ ID NO 1
; LENGTH: 56516
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: promoter
; LOCATION: 1629..1870
; OTHER INFORMATION: Identification method Proscan
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1998..2000
; OTHER INFORMATION: potential start codon
; FEATURE:
; NAME/KEY: exon
; LOCATION: 2001..2216
; OTHER INFORMATION: exon1
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2031..2033
; OTHER INFORMATION: ATG
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 11694..14332
; OTHER INFORMATION: Tyr Phos
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 11930..11947
; OTHER INFORMATION: upstream amplification primer 4-77 SEQ ID42
; FEATURE:
; NAME/KEY: allele
; LOCATION: 12057..12103
; OTHER INFORMATION: polymorphic fragment 4-77 SEQ ID24
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 12339..12358
; OTHER INFORMATION: downstream amplification primer 4-77 SEQ ID51, complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 13547..13564
; OTHER INFORMATION: upstream amplification primer 4-73 SEQ ID64
; FEATURE:
; NAME/KEY: allele
; LOCATION: 13657..13703
; OTHER INFORMATION: polymorphic fragment 4-73 SEQ ID58
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 13962..13981
; OTHER INFORMATION: downstream amplification primer 4-73 SEQ ID67, complement
; FEATURE:
; NAME/KEY: exon
; LOCATION: 18196..18265
; OTHER INFORMATION: exon 2
; FEATURE:
; NAME/KEY: exon
; LOCATION: 23717..23832
; OTHER INFORMATION: exon 3
; FEATURE:
; NAME/KEY: exon
; LOCATION: 25571..25660
; OTHER INFORMATION: exon 4
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 34216..34234
; OTHER INFORMATION: upstream amplification primer 99-217 SEQ ID43
; FEATURE:
; NAME/KEY: allele
; LOCATION: 34469..34515
; OTHER INFORMATION: polymorphic fragment 99-217 SEQ ID25
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 34625..34645
; OTHER INFORMATION: downstream amplification primer 99-217 SEQ ID52, complement

; FEATURE:
; NAME/KEY: exon
; LOCATION: 34669..34759
; OTHER INFORMATION: exon 5
; FEATURE:
; NAME/KEY: exon
; LOCATION: 40688..40846
; OTHER INFORMATION: exon 6
; FEATURE:
; NAME/KEY: exon
; LOCATION: 48070..48193
; OTHER INFORMATION: exon 7
; FEATURE:
; NAME/KEY: exon
; LOCATION: 50182..54523
; OTHER INFORMATION: exon 8
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 51149..51168
; OTHER INFORMATION: upstream amplification primer 4-65 SEQ ID65
; FEATURE:
; NAME/KEY: allele
; LOCATION: 51448..51494
; OTHER INFORMATION: polymorphic fragment 4-65 SEQ ID59
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 51482..51499
; OTHER INFORMATION: downstream amplification primer 4-65 SEQ ID68, complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 51596..51613
; OTHER INFORMATION: upstream amplification primer 4-67 SEQ ID44
; FEATURE:
; NAME/KEY: allele
; LOCATION: 51612..51658
; OTHER INFORMATION: polymorphic fragment 4-67 SEQ ID26
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 51996..52015
; OTHER INFORMATION: downstream amplification primer 4-67 SEQ ID53, complement
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 54445..54450
; OTHER INFORMATION: AATAAA
; US-09-338-907-1

Query Match 48.5%; Score 19.4; DB 4; Length 56516;
Best Local Similarity 79.3%; Pred. No. 71;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 12 cctactctactggcgtgactgggtggg 40
|| ||| || ||||| || |||||
Db 2735 ccaactttaactggcctgaatgctggtggg 2763

RESULT 6
US-09-218-207-1
; Sequence 1, Application US/09218207
; Patent No. 6346331
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Ilya, Chumakov
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: Prostate cancer gene
; FILE REFERENCE: GENSET.018CPI
; CURRENT APPLICATION NUMBER: US/09/218,207
; EARLIER FILING DATE: 1998-12-22
; EARLIER APPLICATION NUMBER: 08/996,306
; EARLIER FILING DATE: 1997-12-22
; EARLIER APPLICATION NUMBER: 60/099,658
; EARLIER FILING DATE: 1998-09-09

```
; NUMBER OF SEQ ID NOS: 578
; SOFTWARE: Patent.pm
; SEQ ID NO 1
;   TYPE: DNA
;   LENGTH: 56516
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: promoter
;   LOCATION: 1629..1870
;   OTHER INFORMATION: identification method Proscan
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1998..2000
;   OTHER INFORMATION: potential start codon
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 2001..2216
;   OTHER INFORMATION: exon1
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 2031..2033
;   OTHER INFORMATION: ATG
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 11694..14332
;   OTHER INFORMATION: Tyr Phos
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 11930..11947
;   OTHER INFORMATION: upstream amplification primer 4-77 SEQ ID42
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 12057..12103
;   OTHER INFORMATION: polymorphic fragment 4-77 SEQ ID24
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 12339..12358
;   OTHER INFORMATION: downstream amplification primer 4-77 SEQ ID51, complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 13547..13564
;   OTHER INFORMATION: upstream amplification primer 4-73 SEQ ID64
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 13657..13703
;   OTHER INFORMATION: polymorphic fragment 4-73 SEQ ID58
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 13962..13981
;   OTHER INFORMATION: downstream amplification primer 4-73 SEQ ID67, complement
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 18196..18265
;   OTHER INFORMATION: exon 2
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 23717..23832
;   OTHER INFORMATION: exon 3
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 25571..25660
;   OTHER INFORMATION: exon 4
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 34216..34234
;   OTHER INFORMATION: upstream amplification primer 99-217 SEQ ID43
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 34469..34515
;   OTHER INFORMATION: polymorphic fragment 99-217 SEQ ID25
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 34625..34645
```

```
; OTHER INFORMATION: downstream amplification primer 99-217 SEQ ID52, complement
; FEATURE:
;   NAME/KEY: exon
;   LOCATION: 34669..34759
;   OTHER INFORMATION: exon 5
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 40688..40846
;   OTHER INFORMATION: exon 6
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 48070..48193
;   OTHER INFORMATION: exon 7
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 50182..54523
;   OTHER INFORMATION: exon 8
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 51149..51168
;   OTHER INFORMATION: upstream amplification primer 4-65 SEQ ID65
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 51448..51494
;   OTHER INFORMATION: polymorphic fragment 4-65 SEQ ID59
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 51482..51499
;   OTHER INFORMATION: downstream amplification primer 4-65 SEQ ID68, complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 51596..51613
;   OTHER INFORMATION: upstream amplification primer 4-67 SEQ ID44
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 51612..51658
;   OTHER INFORMATION: polymorphic fragment 4-67 SEQ ID26
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 51996..52015
;   OTHER INFORMATION: downstream amplification primer 4-67 SEQ ID53, complement
;   FEATURE:
;   NAME/KEY: polyA_signal
;   LOCATION: 54445..54450
;   OTHER INFORMATION: AATAAA
;   US-09-218-207-1

Query Match      48.5%; Score: 19.4; DB 4; Length 56516;
Best Local Similarity 79.3%; Pred. No. 71;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 12 cctactctacctggcgtggactgggtggg 40
    ||||| ||||| ||||| |||||
Db 2735 ccaactttaactggcgttgatgtgtggg 2763

RESULT 7
US-09-338-907-179
; Sequence 179, Application US/09338907
; Patent No. 6265546
; GENERAL INFORMATION:
;   APPLICANT: Cohen, Daniel
;   APPLICANT: Blumenfeld, Marta
;   APPLICANT: Ilya, Chumakov
;   APPLICANT: Bougueleret, Lydie
;   TITLE OF INVENTION: PROSTATE CANCER GENE
;   FILE REFERENCE: GENSET.18CP1CP
;   CURRENT APPLICATION NUMBER: US/09/338,907
;   CURRENT FILING DATE: 1999-06-23
;   EARLIER APPLICATION NUMBER: 08/996,306
;   EARLIER FILING DATE: 1997-12-22
;   EARLIER APPLICATION NUMBER: 60/099,658
```

EARLIER FILING DATE: 1998-09-09
EARLIER APPLICATION NUMBER: 09/218,207
EARLIER FILING DATE: 1998-12-22
NUMBER OF SEQ ID NOS: 578
SOFTWARE: Patent.pm
SEQ ID NO 179
LENGTH: 56520
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: exon
LOCATION: 2001..2216
OTHER INFORMATION: exon1
FEATURE:
NAME/KEY: exon
LOCATION: 18196..18265
OTHER INFORMATION: exon2
FEATURE:
NAME/KEY: exon
LOCATION: 23716..23831
OTHER INFORMATION: exon3
FEATURE:
NAME/KEY: exon
LOCATION: 25570..25659
OTHER INFORMATION: exon4
FEATURE:
NAME/KEY: exon
LOCATION: 34668..34758
OTHER INFORMATION: exon5
FEATURE:
NAME/KEY: exon
LOCATION: 40685..40843
OTHER INFORMATION: exon6
FEATURE:
NAME/KEY: exon
LOCATION: 48067..48190
OTHER INFORMATION: exon7
FEATURE:
NAME/KEY: exon
LOCATION: 50179..54519
OTHER INFORMATION: exon8
FEATURE:
NAME/KEY: polyA_signal
LOCATION: 54493..54498
OTHER INFORMATION: AATAAA
FEATURE:
NAME/KEY: primer_bind
LOCATION: 1991..2008
OTHER INFORMATION: upstream amplification primer 5-63
FEATURE:
NAME/KEY: primer_bind
LOCATION: 2505..2525
OTHER INFORMATION: downstream amplification primer 5-63, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 4091..4111
OTHER INFORMATION: downstream amplification primer 99-622
FEATURE:
NAME/KEY: primer_bind
LOCATION: 4528..4546
OTHER INFORMATION: upstream amplification primer 99-622, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 5475..5495
OTHER INFORMATION: downstream amplification primer 99-621
FEATURE:
NAME/KEY: primer_bind
LOCATION: 5927..5947
OTHER INFORMATION: upstream amplification primer 99-621, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 8127..8144
OTHER INFORMATION: downstream amplification primer 99-619

FEATURE:
NAME/KEY: primer_bind
LOCATION: 8560..8578
OTHER INFORMATION: upstream amplification primer 99-619, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 11622..11639
OTHER INFORMATION: upstream amplification primer 4-76
FEATURE:
NAME/KEY: primer_bind
LOCATION: 12013..12037
OTHER INFORMATION: downstream amplification primer 4-76, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 11930..11947
OTHER INFORMATION: upstream amplification primer 4-77
FEATURE:
NAME/KEY: primer_bind
LOCATION: 12339..12358
OTHER INFORMATION: downstream amplification primer 4-77, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 12915..12932
OTHER INFORMATION: upstream amplification primer 4-71
FEATURE:
NAME/KEY: primer_bind
LOCATION: 13317..13334
OTHER INFORMATION: downstream amplification primer 4-71, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 13216..13233
OTHER INFORMATION: upstream amplification primer 4-72
FEATURE:
NAME/KEY: primer_bind
LOCATION: 13617..13636
OTHER INFORMATION: downstream amplification primer 4-72, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 13547..13564
OTHER INFORMATION: upstream amplification primer 4-73
FEATURE:
NAME/KEY: primer_bind
LOCATION: 13962..13981
OTHER INFORMATION: downstream amplification primer 4-73, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 15994..16011
OTHER INFORMATION: downstream amplification primer 99-610
FEATURE:
NAME/KEY: primer_bind
LOCATION: 16463..16480
OTHER INFORMATION: upstream amplification primer 99-610, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 17304..17324
OTHER INFORMATION: downstream amplification primer 99-609
FEATURE:
NAME/KEY: primer_bind
LOCATION: 17814..17832
OTHER INFORMATION: upstream amplification primer 99-609, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 18008..18027
OTHER INFORMATION: upstream amplification primer 4-90
FEATURE:
NAME/KEY: primer_bind
LOCATION: 18423..18442
OTHER INFORMATION: downstream amplification primer 4-90, complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 18699..18716
OTHER INFORMATION: downstream amplification primer 99-607
FEATURE:


```
; NAME/KEY: primer_bind
; LOCATION: 19164..19182
; OTHER INFORMATION: upstream amplification primer 99-607 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 22589..22609
; OTHER INFORMATION: downstream amplification primer 99-602
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 23111..23129
; OTHER INFORMATION: upstream amplification primer 99-602 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 25098..25118
; OTHER INFORMATION: downstream amplification primer 99-600
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 25657..25674
; OTHER INFORMATION: upstream amplification primer 99-600 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 26537..26557
; OTHER INFORMATION: downstream amplification primer 99-598
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 27022..27040
; OTHER INFORMATION: upstream amplification primer 99-598 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 32262..32281
; OTHER INFORMATION: downstream amplification primer 99-592
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 32823..32841
; OTHER INFORMATION: upstream amplification primer 99-592 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 34215..34233
; OTHER INFORMATION: upstream amplification primer 99-217
; FEATURE:
```

Query Match 48.5%; Score 19.4; DB 4; Length 56520;

Best Local Similarity 79.3%; Pred. No. 71;

Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
Oy 12 cctactctacctggctggactgggtggtg 40
    ||||| ||||| ||||| ||||| |||||
Db 2735 ccaacttaactgggttgatgtgtggtg 2763
```

RESULT 8

US-09-218-207-179

; Sequence 179, Application US/09218207

; Patent No. 6346381

; GENERAL INFORMATION:

; APPLICANT: Cohen, Daniel

; APPLICANT: Blumenfeld, Marta

; APPLICANT: Ilyu, Chumakov

; APPLICANT: Bougueleret, Lydie

; TITLE OF INVENTION: Prostate cancer gene

; FILE REFERENCE: GENSET.018CPI

; CURRENT APPLICATION NUMBER: US/09/218,207

; CURRENT FILING DATE: 1998-12-22

; EARLIER APPLICATION NUMBER: 08/996,306

; EARLIER FILING DATE: 1997-12-22

; EARLIER APPLICATION NUMBER: 60/099,658

; EARLIER FILING DATE: 1998-09-09

; NUMBER OF SEQ ID NOS: 578

; SOFTWARE: Patent.pm

; SEQ ID NO 179

; LENGTH: 56520

; TYPE: DNA

; ORGANISM: Homo sapiens

```
; FEATURE:
; NAME/KEY: exon
; LOCATION: 2001..2216
; OTHER INFORMATION: exon1
; FEATURE:
; NAME/KEY: exon
; LOCATION: 18196..18265
; OTHER INFORMATION: exon2
; FEATURE:
; NAME/KEY: exon
; LOCATION: 23716..23831
; OTHER INFORMATION: exon3
; FEATURE:
; NAME/KEY: exon
; LOCATION: 25570..25659
; OTHER INFORMATION: exon4
; FEATURE:
; NAME/KEY: exon
; LOCATION: 34668..34758
; OTHER INFORMATION: exon5
; FEATURE:
; NAME/KEY: exon
; LOCATION: 40685..40843
; OTHER INFORMATION: exon6
; FEATURE:
; NAME/KEY: exon
; LOCATION: 48067..48190
; OTHER INFORMATION: exon7
; FEATURE:
; NAME/KEY: exon
; LOCATION: 50179..54519
; OTHER INFORMATION: exon8
; NAME/KEY: polyA_signal
; LOCATION: 54493..54498
; OTHER INFORMATION: AATAAA
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 1991..2008
; OTHER INFORMATION: upstream amplification primer 5-63
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 2505..2525
; OTHER INFORMATION: downstream amplification primer 5-63 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 4091..4111
; OTHER INFORMATION: downstream amplification primer 99-622
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 4528..4546
; OTHER INFORMATION: upstream amplification primer 99-622 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 5475..5495
; OTHER INFORMATION: downstream amplification primer 99-621
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 5927..5947
; OTHER INFORMATION: upstream amplification primer 99-621 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 8127..8144
; OTHER INFORMATION: downstream amplification primer 99-619
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 8560..8578
; OTHER INFORMATION: upstream amplification primer 99-619 , complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 11622..11639
; OTHER INFORMATION: upstream amplification primer 4-76
; FEATURE:
```


; APPLICATION NUMBER: 07/946,234
; FILING DATE: 14-SEP-92
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: UIRF89-11A4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (207) 363-0558
; TELEFAX: (207) 363-0528
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2389 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 106..1620
US-08-123-161A-13

Query Match 48.0%; Score 19.2; DB 1; Length 2389;
Best Local Similarity 75.0%; Pred. No. 50;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 3 tcaactgaccctactctacctgggtggactg 34
||||| ||||| || ||||| ||||| |||||
Db 1786 TCACGTGGACCTGCGCGTGTGGGTGGCCTG 1817

RESULT 10
US-08-483-278-13
; Sequence 13, Application US/08483278
; Patent No. 5686073
; GENERAL INFORMATION:
; APPLICANT: Campbell, Kevin P.
; APPLICANT: Ibraghimov, Oxana B.
; APPLICANT: Ervasti, James M.
; APPLICANT: Levellie, Cynthia J.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING DYSTROPHIN-ASSOCIATED
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kevin M. Farrell, P.C.
; STREET: P.O. Box 999
; CITY: York Harbor
; STATE: ME
; COUNTRY: USA
; ZIP: 03911
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,278
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/123,161
; FILING DATE: 16-SEP-93
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: UIRF89-11A5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (207) 363-0558
; TELEFAX: (207) 363-0528
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2389 base pairs
; TYPE: nucleic acid

; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 106..1620
US-08-483-278-13

Query Match 48.0%; Score 19.2; DB 1; Length 2389;
Best Local Similarity 75.0%; Pred. No. 50;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 3 tcaactgaccctactctacctgggtggactg 34
||||| ||||| || ||||| ||||| |||||
Db 1786 TCACGTGGACCTGCGCGTGTGGGTGGCCTG 1817

RESULT 11
US-09-188-930-11
; Sequence 11, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods For Their Use
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 969
; TYPE: DNA
; ORGANISM: mouse
US-09-188-930-11

Query Match 47.5%; Score 19; DB 3; Length 969;
Best Local Similarity 71.4%; Pred. No. 51;
Matches 25; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 6 cttagaccctactctacctgggtggactggtggg 40
||||| ||||| ||||| ||||| ||||| |||||
Db 432 ctagtcccaacttccctggcctggcctggtgtg 466

RESULT 12
US-08-061-465-3
; Sequence 3, Application US/08061465
; Patent No. 5625049
; GENERAL INFORMATION:
; APPLICANT: Monroe, Stephan S.
; APPLICANT: Glass, Roger I.
; APPLICANT: Koopmans, Marion
; APPLICANT: Jiang, Baoming
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING HUMAN ASTROVIRUS
; TITLE OF INVENTION: SEROTYPE II AND USES THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: 127 Peachtree Street, N.E.
; STREET: Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303-1811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

; LOCATION: 189..3833
US-08-486-273A-53

Query Match 47.5%; Score 19; DB 2; Length 4002;
Best Local Similarity 71.4%; Pred. No. 65;
Matches 25; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Oy 3 tcaattgaccctactctacctgggtgactgggt 37
||||| | | | | | | | | | | | | | | |
Db 220 TCACCTCGCTCTTCGGTGCCTGGCAGGCTGGGT 254

RESULT 15
US-08-480-474-53
; Sequence 53, Application US/08480474
; Patent No. 6033865
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie P.
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen W.
; APPLICANT: Lu, Chin-Chun
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR SUBUNITS, DNA ENCODING
; TITLE OF INVENTION: SAME AND USES THEREFOR
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480.474
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-9382B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0999
; TELEFAX: 619-238-0062
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4002 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 189..3833
US-08-480-474-53

Query Match 47.5%; Score 19; DB 3; Length 4002;
Best Local Similarity 71.4%; Pred. No. 65;
Matches 25; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Oy 3 tcaattgaccctactctacctgggtgactgggt 37
||||| | | | | | | | | | | | | | | |
Db 220 TCACCTCGCTCTTCGGTGCCTGGCAGGCTGGGT 254

THIS PAGE BLANK (USPTO)

Result No.	Score	Query Match	Length	DB ID	Description
1	40	100.0	40	22 AAD10601	DNA ligand #8 for
2	29	72.5	40	22 AAD10602	DNA ligand #8 for
C 3	22	55.0	8439	22 AAL37594	Human musculoskele
C 4	22	55.0	9875	22 AAL37595	Human musculoskele
C 5	21.4	55.5	455	22 ABA42775	Human breast cell
C 6	21.4	53.5	455	22 ABA47902	Human breast cell
C 7	21.4	53.5	455	22 ABA53200	Human foetal liver
C 8	21.4	53.5	455	22 ABA65792	Human foetal liver
C 9	21.4	53.5	455	22 ABA32976	Probe #1442 for ge

Polynucleotide ligands useful as anti-viral agents for the treatment of

PT herpesvirus infections e.g. primary or chronic infections with human
XX cytomegalovirus -

PS Claim 8; Page 6; 60pp; English.

XX The invention relates to polynucleotide ligands which bind to
CC herpesvirus, particularly human cytomegalovirus (HCMV) to decrease
CC or block HCMV infection in target cells. These ligands are used as
CC anti-viral agents for the treatment of herpesvirus infections (e.g.
CC lesions and herpetic ulcers) and particularly HCMV associated
CC primary or chronic infections such as pneumonia, gastroenteritis
CC and chorioretinitis. The present sequence is DNA ligand which bind
CC to HCMV to decrease or block HCMV infection in target cells.

XX Sequence 40 BP; 4 A; 14 C; 10 G; 12 T; 0 other;

Query Match 100.0%; Score 40; DB 22; Length 40;

Best Local Similarity 100.0%; Pred. No. 1.3e-06;

Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctatttccaccatattcccttggcccttgggtgtggg 40

||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1 ctatttccaccatattcccttggcccttgggtgtggg 40

RESULT 2

AAD10602

ID AAD10602 standard; DNA; 40 BP.

AC AAD10602;

XX 24-SEP-2001 (first entry)

XX DNA ligand #8 for human cytomegalovirus.

XX Ligand; human cytomegalovirus; HCMV infection; herpetic ulcer;
KW pneumonia; gastroenteritis; chorioretinitis; antiviral agent;
KW herpesvirus infection; lesion; ds.

XX Human cytomegalovirus.

OS Synthetic.

XX Key Location/Qualifiers

PH misc_feature 13

FT /*tag- a
FT /note- "This residue is absent in the sequence
FT shown as SEQ ID NO:16 in the sequence listing
FT of the specification"

FT misc_feature 24..25

FT /*tag- b
FT /note- "This residue is absent in the sequence
FT shown as SEQ ID NO:16 in the sequence listing
FT of the specification"

XX WO200138341-Al.

XX 31-MAY-2001.

XX 21-NOV-2000; 2000WO-US32259.

XX 24-NOV-1999; 99US-0167509.

XX (REGC) UNIV CALIFORNIA.

XX Liu F, Wang J, Jiang H;

XX WPI; 2001-367658/38.

XX Polynucleotide ligands useful as anti-viral agents for the treatment of
PT herpesvirus infections e.g. primary or chronic infections with human
PT cytomegalovirus -

PS Claim 8; Page 6; 60pp; English.

XX The invention relates to polynucleotide ligands which bind to
CC herpesvirus, particularly human cytomegalovirus (HCMV) to decrease
CC or block HCMV infection in target cells. These ligands are used as
CC anti-viral agents for the treatment of herpesvirus infections (e.g.
CC lesions and herpetic ulcers) and particularly HCMV associated
CC primary or chronic infections such as pneumonia, gastroenteritis
CC and chorioretinitis. The present sequence is DNA ligand which bind
CC to HCMV to decrease or block HCMV infection in target cells.

XX Sequence 40 BP; 5 A; 15 C; 6 G; 11 T; 3 other;

Query Match 72.5%; Score 29; DB 22; Length 40;

Best Local Similarity 80.0%; Pred. No. 0.027;

Matches 32; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ctatttccaccatattcccttggcccttgggtgtggg 40

||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1 ctatttccaccatattcccttggcccttgggtgtggg 40

RESULT 3

AAL37594/c

ID AAL37594 standard; DNA; 8439 BP.

XX AAL37594;

XX 08-JAN-2002 (first entry)

XX Human musculoskeletal system related polynucleotide SEQ ID NO 3959.

XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antinflammatory; antiulcer;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein;
KW musculoskeletal system; ds.

XX Homo sapiens.

XX WO20015367-Al.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01338.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0189874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 19-MAY-2000; 2000US-0205515.

XX 07-JUN-2000; 2000US-0209467.

XX 28-JUN-2000; 2000US-0214896.

XX 30-JUN-2000; 2000US-0215135.

XX 07-JUL-2000; 2000US-0216647.

XX 07-JUL-2000; 2000US-0216880.

XX 11-JUL-2000; 2000US-0217487.

XX 14-JUL-2000; 2000US-0218290.

XX 26-JUL-2000; 2000US-0220963.

XX 26-JUL-2000; 2000US-0220964.

XX 14-AUG-2000; 2000US-0224518.

XX 14-AUG-2000; 2000US-0224519.

XX 14-AUG-2000; 2000US-0225213.

XX 14-AUG-2000; 2000US-0225214.

XX 14-AUG-2000; 2000US-0225266.

XX 14-AUG-2000; 2000US-0225267.

XX 14-AUG-2000; 2000US-0225268.

PR	14-AUG-2000;	2000US-02252700;	2000US-02252700;
PR	14-AUG-2000;	2000US-02254477;	2000US-02254477;
PR	14-AUG-2000;	2000US-02255757;	2000US-02255757;
PR	14-AUG-2000;	2000US-02255758;	2000US-02255758;
PR	14-AUG-2000;	2000US-02255759;	2000US-02255759;
PR	18-AUG-2000;	2000US-02262579;	2000US-02262579;
PR	22-AUG-2000;	2000US-02266861;	2000US-02266861;
PR	22-AUG-2000;	2000US-02266868;	2000US-02266868;
PR	23-AUG-2000;	2000US-02271827;	2000US-02271827;
PR	23-AUG-2000;	2000US-02277009;	2000US-02277009;
PR	30-AUG-2000;	2000US-0228924;	2000US-0228924;
PR	01-SEP-2000;	2000US-02292877;	2000US-02292877;
PR	01-SEP-2000;	2000US-02293433;	2000US-02293433;
PR	01-SEP-2000;	2000US-02293444;	2000US-02293444;
PR	01-SEP-2000;	2000US-02293455;	2000US-02293455;
PR	05-SEP-2000;	2000US-02295099;	2000US-02295099;
PR	05-SEP-2000;	2000US-02295113;	2000US-02295113;
PR	06-SEP-2000;	2000US-02304337;	2000US-02304337;
PR	06-SEP-2000;	2000US-02304339;	2000US-02304339;
PR	08-SEP-2000;	2000US-02311242;	2000US-02311242;
PR	08-SEP-2000;	2000US-02311243;	2000US-02311243;
PR	08-SEP-2000;	2000US-02311244;	2000US-02311244;
PR	08-SEP-2000;	2000US-02311413;	2000US-02311413;
PR	08-SEP-2000;	2000US-02311414;	2000US-02311414;
PR	08-SEP-2000;	2000US-02320800;	2000US-02320800;
PR	08-SEP-2000;	2000US-02320811;	2000US-02320811;
PR	12-SEP-2000;	2000US-02319688;	2000US-02319688;
PR	14-SEP-2000;	2000US-02323397;	2000US-02323397;
PR	14-SEP-2000;	2000US-02330633;	2000US-02330633;
PR	14-SEP-2000;	2000US-02330634;	2000US-02330634;
PR	14-SEP-2000;	2000US-02330655;	2000US-02330655;
PR	21-SEP-2000;	2000US-02342233;	2000US-02342233;
PR	21-SEP-2000;	2000US-02342777;	2000US-02342777;
PR	25-SEP-2000;	2000US-02349597;	2000US-02349597;
PR	25-SEP-2000;	2000US-02349998;	2000US-02349998;
PR	26-SEP-2000;	2000US-02354584;	2000US-02354584;
PR	27-SEP-2000;	2000US-02358334;	2000US-02358334;
PR	27-SEP-2000;	2000US-02358336;	2000US-02358336;
PR	29-SEP-2000;	2000US-02363237;	2000US-02363237;
PR	29-SEP-2000;	2000US-02363677;	2000US-02363677;
PR	29-SEP-2000;	2000US-02363688;	2000US-02363688;
PR	29-SEP-2000;	2000US-02363699;	2000US-02363699;
PR	29-SEP-2000;	2000US-02363700;	2000US-02363700;
PR	02-OCT-2000;	2000US-02366802;	2000US-02366802;
PR	02-OCT-2000;	2000US-02370337;	2000US-02370337;
PR	02-OCT-2000;	2000US-02370338;	2000US-02370338;
PR	02-OCT-2000;	2000US-02370339;	2000US-02370339;
PR	02-OCT-2000;	2000US-02370400;	2000US-02370400;
PR	13-OCT-2000;	2000US-02399337;	2000US-02399337;
PR	13-OCT-2000;	2000US-02399337;	2000US-02399337;
PR	20-OCT-2000;	2000US-02409600;	2000US-02409600;
PR	20-OCT-2000;	2000US-02411221;	2000US-02411221;
PR	20-OCT-2000;	2000US-02417855;	2000US-02417855;
PR	20-OCT-2000;	2000US-02417856;	2000US-02417856;
PR	01-NOV-2000;	2000US-02446177;	2000US-02446177;
PR	08-NOV-2000;	2000US-02464774;	2000US-02464774;
PR	08-NOV-2000;	2000US-02464775;	2000US-02464775;
PR	08-NOV-2000;	2000US-02464776;	2000US-02464776;
PR	08-NOV-2000;	2000US-02464777;	2000US-02464777;
PR	08-NOV-2000;	2000US-02464778;	2000US-02464778;
PR	08-NOV-2000;	2000US-02465233;	2000US-02465233;
PR	08-NOV-2000;	2000US-02465234;	2000US-02465234;
PR	08-NOV-2000;	2000US-02465235;	2000US-02465235;
PR	08-NOV-2000;	2000US-02465236;	2000US-02465236;
PR	08-NOV-2000;	2000US-02465237;	2000US-02465237;

PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	08-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	03-DEC-2000;	2000US-0251030.
PR	03-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.

Query Match 55.0%; Score 22; DB 22; Length 8439;
Best Local Similarity 83.3%; Pred. No. 33;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 7 cccaccatcccttggcccttgggtg 36
||||| ||||| ||||| |||
Db 5444 CCCACCACATCCCTGGGCCCTGAGATG 5415

RESULT 4
AAL37595/c
ID AAL37595 standard; DNA; 9875 BP.
XX
AC AAL37595;
XX
DT 08-JAN-2002 (first entry)
DE Human musculoskeletal system related polynucleotide SEQ ID NO 3960.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
DE anti-allergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein;
KW musculoskeletal system; ds.

XX Homo sapiens.
XX
PN WO20015367-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01338.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.


```
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00669.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver -
XX Claim 4; SEQ ID NO 14097; 639pp + sequence listing; English.
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX fetal liver. The present sequence is a single exon nucleic acid
XX probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 455 BP; 104 A; 134 C; 129 G; 88 T; 0 other;

Query Match 53.5%; Score 21.4; DB 22; Length 455;
Best Local Similarity 71.8%; Pred. No. 37;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Oy 2 tatttccaccatcccttggcccttggtgtgg 40
Db | ||||| ||||| || ||||| || ||||| ||
90 TGTTCCTCCACCCAAATTCCTAGTCCACTGGGGGAGAG 52

RESULT 9
ABA22976/c
ID ABA22976 standard; DNA; 455 BP.
XX AC ABA22976;
XX DT 23-JAN-2002 (first entry)
XX DE Probe #1442 for gene expression analysis in human heart cell sample.
XX KW Human; gene expression; heart; microarray; vascular system; probe;
XX KW cardiovascular disease; hypertension; cardiac arrhythmia;
XX KW congenital heart disease; ss.
XX OS Homo sapiens.
XX PN WO200157274-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00666.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver -
XX Claim 4; SEQ ID NO 14097; 639pp + sequence listing; English.
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX fetal liver. The present sequence is a single exon nucleic acid
XX probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 455 BP; 104 A; 134 C; 129 G; 88 T; 0 other;

Query Match 53.5%; Score 21.4; DB 22; Length 455;
Best Local Similarity 71.8%; Pred. No. 37;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Oy 2 tatttccaccatcccttggcccttggtgtgg 40
Db | ||||| ||||| || ||||| || ||||| ||
90 TGTTCCTCCACCCAAATTCCTAGTCCACTGGGGGAGAG 52

RESULT 10
ABA32881/c
ID ABA32881 standard; DNA; 455 BP.
XX AC ABA32881;
XX DT 23-JAN-2002 (first entry)
XX DE Probe #11347 for gene expression analysis in human heart cell sample.
XX KW Human; gene expression; heart; microarray; vascular system; probe;
XX KW cardiovascular disease; hypertension; cardiac arrhythmia;
XX KW congenital heart disease; ss.
XX OS Homo sapiens.
XX PN WO200157274-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00666.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488899/53.
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts -
XX Claim 1; SEQ ID NO 1442; 530pp; English.
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart. The
XX present sequence is one such probe. The probes may be used for
XX predicting, measuring and displaying gene expression in samples derived
XX from the human heart via microarrays. By measuring gene expression, the
XX probes are useful for predicting, diagnosing, grading, staging,
XX monitoring and prognosing diseases of the human heart and vascular system
XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
XX congenital heart disease.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 455 BP; 104 A; 134 C; 129 G; 88 T; 0 other;
```

DR WPI; 2001-488899/53.
XX Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
PS Claim 4; SEQ ID NO 11347; 530pp; English.
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 455 BP; 104 A; 134 C; 129 G; 88 T; 0 other;

Query Match 53.5%; Score 21.4; DB 22; Length 455;
Best Local Similarity 71.8%; Pred. No. 37;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 2 tatttccacccatcccccttggtggcccttggtgtggg 40
I|||||I|||||I|||||I|||||I|||||I|||||I
DB 90 TGTTCACCCAAATTCCTAGCTCCACTGGGGGAGAG 52

RESULT 11
AAK01450/c
ID AAK01450 standard; DNA: 455 BP.
XX
AC AAK01450;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 1441.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00667.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX
XX Example 4; SEQ ID NO: 1441; 650pp + Sequence Listing; English.

CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
SQ Sequence 455 BP; 104 A; 134 C; 129 G; 88 T; 0 other;

Query Match 53.5%; Score 21.4; DB 22; Length 455;
Best Local Similarity 71.8%; Pred. No. 37;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 2 tatttccacccatcccccttggtggcccttggtgtggg 40
I|||||I|||||I|||||I|||||I|||||I|||||I
DB 90 TGTTCACCCAAATTCCTAGCTCCACTGGGGGAGAG 52

RESULT 12
AAK14197/c
ID AAK14197 standard; DNA: 455 BP.
XX
AC AAK14197;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 14188.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00667.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX
XX Example 4; SEQ ID NO: 14188; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is one of the probes of the
XX invention.
XX
SQ Sequence 455 BP; 104 A; 134 C; 129 G; 88 T; 0 other;

Query Match 53.5%; Score 21.4; DB 22; Length 455;

Best Local Similarity 71.8%; Pred. No. 37; Mismatches 0; Indels 11; Gaps 0; Matches 28; Conservative 0; Mismatches 0; Indels 11; Gaps 0;

Qy 2 tatttccaccaccatcccccttgggccccttgggtgtggg 40
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 90 TGTTCCTCCACCAATTCCTAGCTCCACTGGGGGAGAG 52

RESULT 13
AAK26895/c
ID AAK26895 standard; DNA; 455 BP.
XX
AC AAK26895;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 1452.
XX
KW Human: bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
PS Example 4; SEQ ID NO: 1452; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
SQ Sequence 455 BP; 104 A; 134 C; 129 G; 88 T; 0 other;

Query Match 53.5%; Score 21.4; DB 22; Length 455;
Best Local Similarity 71.8%; Pred. No. 37;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 2 tatttccaccaccatcccccttgggccccttgggtgtggg 40
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 90 TGTTCCTCCACCAATTCCTAGCTCCACTGGGGGAGAG 52

RESULT 14
AAK39928/c
ID AAK39928 standard; DNA; 455 BP.
XX
AC AAK39928;
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #1457 for gene expression analysis in human cervical cell sample.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.

DT 06-NOV-2001 (first entry)
DE Human bone marrow expressed single exon probe SEQ ID NO: 14485.
XX
KW Human: bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
PS Example 4; SEQ ID NO: 14485; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
SQ Sequence 455 BP; 104 A; 134 C; 129 G; 88 T; 0 other;

Query Match 53.5%; Score 21.4; DB 22; Length 455;
Best Local Similarity 71.8%; Pred. No. 37;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 2 tatttccaccaccatcccccttgggccccttgggtgtggg 40
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 90 TGTTCCTCCACCAATTCCTAGCTCCACTGGGGGAGAG 52

RESULT 15
AA11524/c
ID AA11524 standard; DNA; 455 BP.
XX
AC AA11524;
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #1457 for gene expression analysis in human cervical cell sample.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.

```
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488901/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
XX Claim 25; SEQ ID No 1457; 487pp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 455 BP; 104 A; 134 C; 129 G; 88 T; 0 other;
```

Query Match 53.5%; Score 21.4; DB 22; Length 455;
Best Local Similarity 71.8%; Pred. No. 37;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 2 tatttccaccatattcccttggcccttgggtgtggg 40
| | | | | | | | | | | | | | | | | | | | | |
Db 90 TGTTCACCCCAATTCCTAGCTCCACTGGGGGAG 52

Search completed: June 4, 2002, 17:56:22
Job time: 7110 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 17:56:14 ; Search time 521.5 Seconds
(without alignments)
131.690 Million cell updates/sec

Title: US-09-721-543A-14
Perfect score: 40
Sequence: 1 cataactgacctactctactctggtgactggtggg 40

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:.*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	100.0	40	22	AA10600
2	22.6	56.5	474	22	AA11954
3	22.6	56.5	650	22	AA120844
4	22.2	55.5	2544	22	AA117822
5	21.2	53.0	119	22	AAH33899
6	20.8	52.0	2687	22	AA17508
7	20.8	52.0	3563	22	AA17517
8	20.8	52.0	24813	22	AAK70879
9	20.6	51.5	2113	22	AA126576
					DNA ligand #6 for
					Human breast cancer
					Human breast cancer
					Human CDNA sequenc
					Human colon cancer
					Human taste recept
					Human taste recept
					Human immune/haema
					Human breast cancer

10	20.4	51.0	615	21	AAF04663	Fusarium venenatum
11	20.4	51.0	23474	23	AAS58511	Propionibacterium
12	20.4	51.0	24908	22	AAK81665	Human immune/haema
13	20.2	50.5	359	21	AAZ58767	Human huntingtin-i
14	20.2	50.5	601	24	AA561792	Human small cell ca
15	20.2	50.5	1164	18	AAAT67186	Huntingtin interac
16	20.2	50.5	1164	21	AAZ58745	Human huntingtin-i
17	20.2	50.5	4796	18	AAAT67187	Huntingtin interac
18	20.2	50.5	4796	21	AAZ58746	Human huntingtin-i
19	20.2	50.0	466	21	AAZ26742	Human secreted pro
20	20.2	50.0	531	22	AAK64603	Human immune/haema
21	20.2	50.0	3210	23	ABL07885	Drosophila melanog
22	20.2	50.0	4377	23	ABL07884	Drosophila melanog
23	19.8	49.5	509	23	AA567978	DNA encoding novel
24	19.8	49.5	1280	23	AA581491	DNA encoding novel
25	19.8	49.5	1371	23	ABL14959	Drosophila melanog
26	19.8	49.5	1713	22	ABL36883	Human musculoskele
27	19.8	49.5	1830	22	AAH14834	Human cDNA sequenc
28	19.8	49.5	1849	22	AAH64889	Human secreted pro
29	19.8	49.5	2086	23	AA584430	DNA encoding novel
30	19.8	49.5	2497	22	AA158171	Human polynucleoti
31	19.8	49.5	2511	22	ABA09097	Human HSPC264 homo
32	19.8	49.5	3969	18	AAAT62655	DNA encoding prote
33	19.8	49.5	4319	22	AA105694	Human reproductive
34	19.8	49.5	4813	23	ABL14958	Drosophila melanog
35	19.8	49.5	5351	22	AA531466	Human DNA for a no
36	19.8	49.5	14176	22	AA526670	Human genomic DNA
37	19.8	49.5	32249	22	ABA17155	Human nervous syst
38	19.6	49.0	718	11	AAQ03934	Ovine growth hormo
39	19.6	49.0	817	10	AAQ90301	Goat growth hormo
40	19.6	49.0	1717	21	AAAC78069	Human cancer assoc
41	19.6	49.0	2064	21	AA665124	Porcine BAC-PGF2-
42	19.6	49.0	2230	9	AAH81778	Sequence of porcin
43	19.6	49.0	2638	22	ABA21192	Human nervous syst
44	19.6	49.0	3211	22	AAH34203	Human colon cancer
45	19.6	49.0	3820	20	AAZ25349	Rat DRM nucleotide

ALIGNMENTS

RESULT 1	
AA10600	
ID	AA10600 standard; DNA; 40 BP.
XX	
AC	AA10600;
XX	
DT	24-SEP-2001 (first entry)
XX	
DE	DNA ligand #6 for human cytomegalovirus.
XX	
KW	Ligand: human cytomegalovirus; HCMV infection; herpetic ulcer;
KW	pneumonia; gastroenteritis; chorioretinitis; antiviral agent;
KW	herpesvirus infection; lesion; ds.
XX	
OS	Human cytomegalovirus.
OS	Synthetic.
XX	
PN	WO200138341-A1.
XX	
PD	31-MAY-2001.
XX	
PF	21-NOV-2000; 2000WO-US32259.
XX	
PR	24-NOV-1999; 99US-0167509.
XX	
PA	(REGC) UNIV CALIFORNIA.
XX	
PI	Liu F, Wang J, Jiang H;
XX	
DR	WPI; 2001-367658/38.
XX	
PT	Polynucleotide ligands useful as anti-viral agents for the treatment of

RESULT	4
AAH17822	
ID	AAH17822 standard; cDNA; 2544 BP.
XX	
AC	AAH17822;
XX	
DT	26-JUN-2001 (first entry)
XX	
DE	Human cDNA sequence SEQ ID NO:17492.
XX	
KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX	
OS	Homo sapiens.
XX	
PN	EP1074617-A2.
XX	
PD	07-FEB-2001.
XX	
PF	28-JUL-2000; 2000EP-0116126.
XX	
PR	29-JUL-1999; 99JP-0248036.
PR	27-AUG-1999; 99JP-0300253.
PR	11-JAN-2000; 2000JP-0118776.
PR	02-MAY-2000; 2000JP-0183767.
PR	09-JUN-2000; 2000JP-0241899.
XX	
PA	(HELI-) HELIX RES INST.
XX	
PI	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX	
DR	WPI; 2001-318749/34.
XX	
PT	Primer sets for synthesizing polynucleotides, particularly the 5602
PT	full-length cDNAs defined in the specification, and for the detection
PT	and/or diagnosis of the abnormality of the proteins encoded by the
PT	full-length cDNAs -
XX	
PS	Claim 8; SEQ ID 17492; 2537pp + CD ROM; English.
XX	
CC	The present invention describes primer sets for synthesising 5602
CC	full-length cDNAs defined in the specification. Where a primer set
CC	comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC	to the complementary strand of a polynucleotide which comprises one of
CC	the 5602 nucleotide sequences defined in the specification, where the
CC	oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC	of an oligonucleotide comprising a sequence complementary to the
CC	complementary strand of a polynucleotide which comprises a 5'-end
CC	sequence and an oligonucleotide comprising a sequence complementary to a
CC	polynucleotide which comprises a 3'-end sequence, where the
CC	oligonucleotide comprises at least 15 nucleotides and the combination of
CC	the 5'-end sequence/3'-end sequence is selected from those defined in
CC	the specification. The primer sets can be used in antisense therapy and
CC	in gene therapy. The primers are useful for synthesising polynucleotides,
CC	particularly full-length cDNAs. The primers are also useful for the
CC	detection and/or diagnosis of the abnormality of the proteins encoded by
CC	the full-length cDNAs. The primers allow obtaining of the full-length
CC	cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC	AAH13633 to AAH18742 represent human cDNA sequences; AA892446 to
CC	AA893893 represent human amino acid sequences; and AAH13629 to AAH13632
CC	represent oligonucleotides, all of which are used in the exemplification
CC	of the present invention.
XX	
SQ	Sequence 2544 BP; 467 A; 775 C; 740 G; 562 T; 0 other;
Query Match	55.5%; Score 22.2; DB.22; Length 2544;
Best Local Similarity	77.1%; Pred. No. 22;
Matches	27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
Qy	6 cttgaccactactactcctggcgactgggctggg 40
Db	
	1138 cttgaacctctgcctcctgcctgcctattggg 1172

RESULT	5
AAH33899	
ID	AAH33899 standard; cDNA; 119 BP.
XX	
AC	AAH33899;
XX	
DT	03-SEP-2001 (first entry)
XX	
DE	Human colon cancer antigen encoding cDNA SEQ ID NO:955.
XX	
KW	Human; colon cancer; colon cancer antigen; diagnosis; detection; colorectal carcinoma; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200122920-A2.
XX	
PD	05-APR-2001.
XX	
PF	28-SEP-2000; 2000WO-US26524.
XX	
PR	29-SEP-1999; 99US-0157137.
XX	
PR	03-NOV-1999; 99US-0163280.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
PI	Ruben SM, Barash SC, Birse CE, Rosen CA;
XX	
DR	WPI; 2001-235357/24.
XX	
DR	P-PDSB; AAG74468.
XX	
PT	Nucleic acids encoding 4277 human colon cancer-associated polypeptides, useful for preventing, diagnosing and/or treating colorectal cancers -
XX	
PS	Claim 1; Page 2853; 9803pp; English.
XX	
CC	AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon cancer-associated nucleic acid molecules (N) and proteins (P), where the proteins are collectively known as colon cancer antigens. The colon cancer antigens have cytostatic activity and can be used in gene therapy and vaccine production. N and P may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate P expression. For example, N and P may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of P by expressing P. Inactive proteins or to supplement the patients own production of P. Additionally, N may be used to produce the colon cancer-associated Ps, by inserting the nucleic acids into a host cell and culturing the cell to express the proteins. N and P can be used in the prevention, diagnosis and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204 and AAH77789 represent sequences used in the exemplification of the present invention.
CC	N.B. Pages 666 to 682 and page 7053 of the sequence listing were missing at time of publication, meaning no sequences are present for SEQ ID NO:1027 to 1052, 7921 and 7922.
XX	
SQ	Sequence 119 BP; 10 A; 47 C; 38 G; 23 T; 1 other;
Query Match	53.0%; Score 21.2; DB 22; Length 119;
Best Local Similarity	72.2%; Fred. No. 36;
Matches	26; Conservative 1; Mismatches 9; Indels 0; Gaps 0
Qy	5 acttgaccctactctacctggcgctgacctgggtggg 40 :
Dd	39 accctgaacctctgcmtcgtggcgctcccccactggg 74
RESULT	6
AAD17508/c	
ID	AAD17508 standard; DNA; 2687 BP.

Qy 6 ctgagccctactctacctgggctggagactgggtggg 40
||||| ||| ||| |||||||| ||| |||||
Db 1138 ctgaacctctgcctgggctggcctgattggg 1172


```
FT      /*tag= k
FT      /number= 5
FT      intron 2297..2422
FT      /*tag= l
FT      /number= 6
FT      exon    2423..3381
FT      /*tag= m
FT      /number= 6
FT      intron  3382..3563
FT      /*tag= n
FT      /number= 7
FT      /cons_splice= (5'site:NO, 3'site:NO)
XX      WO200166563-A2.
XX
XX      13-SEP-2001.
XX
XX      07-MAR-2001; 2001WO-US07265.
XX
XX      07-MAR-2000; 2000US-0187546.
XX      07-APR-2000; 2000US-0195536.
XX      06-JUN-2000; 2000US-0209840.
XX      23-JUN-2000; 2000US-0214213.
XX      17-AUG-2000; 2000US-0226448.
XX      03-JAN-2001; 2001US-0259227.
XX
XX      (SENO-) SENOMYX INC.
XX
XX      Adler JE, Zozulya S, O'Connell SM, Li X, Staszewski L;
XX
XX      WPI: 2001-582267/65.
XX      P-PSDB; AAE10366.
XX
XX      New mammalian taste-cell-specific G protein-coupled receptor
XX      polypeptides for identifying compounds that modulate taste signaling
XX      are useful in food, to modulate the sweet taste of foods or drugs -
XX
XX      Claim 1; Page 71-72; 119pp; English.
XX
XX      The invention relates to mammalian taste-cell-specific G protein-coupled
XX      receptors, T1R and their corresponding cDNA molecules. Taste receptors,
XX      T1R are useful for screening compounds which are used to activate or
XX      modulate chemosensory transduction, such as taste sensation. The
XX      identification and isolation of novel taste receptors and taste
XX      signalling molecules allow for new methods of chemical and genetic
XX      modulation of taste transduction pathways. The taste modulating
XX      compounds are useful in pharmaceuticals and food industries to improve
XX      the taste of a variety of consumer products, or to block undesirable
XX      tastes, e.g., in certain pharmaceuticals. T1R.s are also useful in
XX      biochemical assay for identifying tastant (T1R) ligands having binding
XX      specificity for T1R involved in taste signalling. The present sequence is
XX      human taste-cell-specific G protein-coupled receptor, hT1R3 full-length
XX      genomic DNA. Human T1R3 gene is localised on chromosome 1p36.2-1p36.33.
XX
XX      Sequence 3563 BP; 545 A; 1214 C; 1166 G; 638 T; 0 other;
XX
XX      Query Match      52.0%; Score 20.8; DB 22; Length 3563;
XX      Best Local Similarity 70.0%; Pred. No. 83;
XX      Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
XX
OY      1 catcacttgaccctactctactacgtggctggactgggtggg 40
Db      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      925 CATCTTCTGTGACACAGGGCTCCCGAGGCGAGGCTGGTGGG 886

RESULT      8
AAK70879
ID      AAK70879 standard; DNA; 24813 BP.
XX
XX      AAK70879;
XX
XX      06-NOV-2001 (first entry)
XX
```

Human immune/haematopoietic antigen genomic sequence SEQ ID NO:25691.
Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
cystostatic; gene therapy; vaccine; metastasis; ds.
Homo sapiens.
WO200157182-A2.
09-AUG-2001.
17-JAN-2001; 2001WO-US01354.
31-JAN-2000; 2000US-0179065.
04-FEB-2000; 2000US-0180628.
24-FEB-2000; 2000US-0184664.
02-MAR-2000; 2000US-0186350.
16-MAR-2000; 2000US-0189874.
17-MAR-2000; 2000US-0190076.
18-APR-2000; 2000US-0198123.
19-MAY-2000; 2000US-0205515.
07-JUN-2000; 2000US-0209467.
28-JUN-2000; 2000US-0214886.
30-JUN-2000; 2000US-0215135.
07-JUL-2000; 2000US-0216647.
07-JUL-2000; 2000US-0216880.
11-JUL-2000; 2000US-0217487.
11-JUL-2000; 2000US-0217496.
14-JUL-2000; 2000US-0218290.
26-JUL-2000; 2000US-0220963.
26-JUL-2000; 2000US-0220964.
14-AUG-2000; 2000US-0224518.
14-AUG-2000; 2000US-0224519.
14-AUG-2000; 2000US-0225213.
14-AUG-2000; 2000US-0225214.
14-AUG-2000; 2000US-0225266.
14-AUG-2000; 2000US-0225267.
14-AUG-2000; 2000US-0225268.
14-AUG-2000; 2000US-0225270.
14-AUG-2000; 2000US-0225447.
14-AUG-2000; 2000US-0225757.
14-AUG-2000; 2000US-0225758.
14-AUG-2000; 2000US-0225759.
18-AUG-2000; 2000US-0226279.
22-AUG-2000; 2000US-0226681.
22-AUG-2000; 2000US-0226688.
22-AUG-2000; 2000US-0227182.
23-AUG-2000; 2000US-0227009.
30-AUG-2000; 2000US-0228924.
01-SEP-2000; 2000US-0229287.
01-SEP-2000; 2000US-0229343.
01-SEP-2000; 2000US-0229344.
01-SEP-2000; 2000US-0229345.
05-SEP-2000; 2000US-0229509.
05-SEP-2000; 2000US-0229513.
06-SEP-2000; 2000US-0230437.
06-SEP-2000; 2000US-0230438.
08-SEP-2000; 2000US-0231242.
08-SEP-2000; 2000US-0231243.
08-SEP-2000; 2000US-0231244.
08-SEP-2000; 2000US-0231413.
08-SEP-2000; 2000US-0231414.
08-SEP-2000; 2000US-0232080.
08-SEP-2000; 2000US-0232081.
12-SEP-2000; 2000US-0231968.
14-SEP-2000; 2000US-0232397.
14-SEP-2000; 2000US-0232398.
14-SEP-2000; 2000US-0232399.
14-SEP-2000; 2000US-0232400.
14-SEP-2000; 2000US-0232401.
14-SEP-2000; 2000US-0233063.
14-SEP-2000; 2000US-0233064.
14-SEP-2000; 2000US-0233064.

```
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234997.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235834.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251398.
PR 06-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.

PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis.
XX
XX Disclosure; SEQ ID NO 25691; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
XX amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patient's own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting
XX the nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/haematopoietic-related diseases, especially
XX cancers and cancer metastases of haematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/haematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAK82169
XX represent sequences used in the exemplification of the present invention.
XX
XX Sequence 24813 BP; 5290 A; 6858 C; 7656 G; 5009 T; 0 other;

Query Match 52.0%; Score 20.8; DB 22; Length 24813;
Best Local Similarity 78.1%; Pred. No. 1.1e+02;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 8 tgaccctactctacgtggctggactgggtgg 39
    ||||| |||| ||| ||||| |||||
Db 8205 tgacctgtctctctgtggcctggcgtgtgg 8236

RESULT 9
AAL26576/c
ID AAL26576 standard; cDNA; 2113 BP.
XX
XX AAL26576;
XX
XX 07-DEC-2001 (first entry)
XX
XX Human breast cancer expressed polynucleotide 19033.
XX
XX Human; breast cancer; cell marker; cytostatic; ss.
XX
XX Homo sapiens.
XX
XX WO200151628-A2.
XX
XX 19-JUL-2001.
XX
XX 10-JAN-2001; 2001WO-US00798.
XX
XX 14-JAN-2000; 2000US-0176077.
XX 14-MAR-2000; 2000US-0189167.
XX 24-MAR-2000; 2000US-0192099.
XX 29-MAR-2000; 2000US-0193480.
XX 15-JUN-2000; 2000US-0205230.
XX 09-JUN-2000; 2000US-0211315.
```


[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 17:46:31 ; Search time 2137.99 Seconds
(without alignments)
391.518 Million cell updates/sec

Title: US-09-721-543A-14
Perfect score: 40
Sequence: 1 catcaactgaccctactctactctggtgactgggtggg 40

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*

- 1: gb_da:*
- 2: gb_hgt:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pi:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vl:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vl:*
- 30: em_hgt_hum:*
- 31: em_hgt_inv:*
- 32: em_hgt_other:*
- 33: em_hgtgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description

C	1	24.2	60.5	125817	2	AC087325	AC087325 Trypanoso
C	2	23.8	59.5	40649	9	AC003111	AC003111 Human DNA
C	3	23.8	59.5	42497	9	AC004798	AC004798 Homo sapi
C	4	22.8	57.0	45420	2	AC109521	AC109521 Mus muscu
C	5	22.8	57.0	157076	9	AC019031	AC019031 Homo sapi
C	6	22.8	57.0	212671	2	AC084849	AC084849 Homo sapi
C	7	22.8	57.0	213611	2	AC091573	AC091573 Homo sapi
C	8	22.8	57.0	213621	9	AF276759	AF276759 Homo sapi
C	9	22.6	56.5	112188	2	AC098541	AC098541 Rattus no
C	10	22.6	56.5	162044	9	AL390763	AL390763 Human DNA
C	11	22.6	56.5	163357	2	AC023276	AC023276 Homo sapi
C	12	22.6	56.5	172385	2	AL591127	AL591127 Mus muscu
C	13	22.6	56.5	178660	9	HSG1150	AJ006996 Homo sapi
C	14	22.6	56.5	340000	9	AP001700	AP001700 Homo sapi
C	15	22.4	56.0	333	11	G42160	G42160 ps701068 KW
C	16	22.4	56.0	68118	2	AC011113	AC011113 Homo sapi
C	17	22.4	56.0	139515	9	HS38C16	AL035604 Human DNA
C	18	22.4	56.0	150147	9	AC002543	AC002543 Homo sapi
C	19	22.4	56.0	160543	9	AC010271	AC010271 Homo sapi
C	20	22.4	56.0	164652	2	AC021625	AC021625 Homo sapi
C	21	22.4	56.0	206717	2	AL645535	AL645535 Mus muscu
C	22	22.4	56.0	222232	2	AL645963	AL645963 Mus muscu
C	23	22.2	55.5	2544	9	AK023107	AK023107 Homo sapi
C	24	22.2	55.5	142645	9	AC008481	AC008481 Homo sapi
C	25	22.2	55.5	172705	9	AC013372	AC013372 Homo sapi
C	26	22.2	55.5	176894	2	AL355574	AL355574 Homo sapi
C	27	22.2	55.5	181890	2	AL355987	AL355987 Homo sapi
C	28	22.2	55.5	187883	9	AC008268	AC008268 Homo sapi
C	29	22.2	55.5	248846	2	AC090008	AC090008 Mus muscu
C	30	22.2	55.0	767	8	AF168146	AF168146 Calliervo
C	31	22.2	55.0	109255	2	AC098163	AC098163 Rattus no
C	32	22.2	55.0	167118	2	AC098129	AC098129 Rattus no
C	33	22.2	55.0	181807	9	AL356267	AL356267 Human DNA
C	34	21.6	54.0	1511	8	ZMANT2NU	X59086 Z. mays MANT
C	35	21.6	54.0	2010	14	AF105468	AF105468 Feline ca
C	36	21.6	54.0	104509	2	AC095464	AC095464 Rattus no
C	37	21.6	54.0	110000	2	AC103319	AC103319 Rattus no
C	38	21.6	54.0	115364	2	AC094857	AC094857 Rattus no
C	39	21.6	54.0	173342	9	AC023796	AC023796 Homo sapi
C	40	21.6	54.0	185306	2	AC013570	AC013570 Homo sapi
C	41	21.6	54.0	209345	2	AC106686	AC106686 Homo sapi
C	42	21.4	53.5	9685	1	STY242516	AJ242516 Salmonell
C	43	21.4	53.5	22286	1	AE008700	AE008700 Salmonell
C	44	21.4	53.5	75684	9	AL512843	AL512843 Human DNA
C	45	21.4	53.5	119939	2	AC098001	AC098001 Rattus no

ALIGNMENTS

RESULT	1
AC087325/c	
LOCUS	125817 bp DNA linear HTG 17-JUL-2001
DEFINITION	Trypanosoma brucei chromosome IV clone RPCI93-29M18, *** SEQUENCING IN PROGRESS ***, 1 ordered pieces.
ACCESSION	AC087325
VERSION	AC087325.8 GI:14787198
KEYWORDS	HTG: HTGS_PHASE2.
SOURCE	Trypanosoma brucei.
ORGANISM	Trypanosoma brucei
REFERENCE	1 (bases 1 to 125817) El-Sayed,N.M., Ghedin,E., Song,J., Larkin,C., Wanless,D., Jones,K., Peterson,J., Hou,L., Zhao,H., Mason,T., Millican,J., Pa,G., Van Aken,S., Uterback,T., Khalak,H.G., Gerard,C., Leech,V., Ullu,E., Melville,S., White,O., Adams,M.D., Donelson,J.E. and Fraser,C.M. Trypanosoma brucei GUTat10.1 RPCI93-29M18 BAC genomic sequence Unpublished
JOURNAL	2 (bases 1 to 125817)
AUTHORS	El-Sayed,N.M., Khalak,H. and Adams,M.D.
TITLE	Direct Submission
JOURNAL	Submitted (25-DEC-2000) The Institute for Genomic Research, 9712

COMMENT

Medical Center Dr, Rockville, MD 20850, USA
 On Jul 17, 2001 this sequence version replaced gi:13277449.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as runs of N, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submittor.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 125817: contig of 125817 bp in length.

FEATURES

source

1..125817
 /organism="Trypanosoma brucei"
 /isolate="GUTat10.1"
 /db_xref="taxon:5691"
 /chromosome="IV"
 /clone="RPC193-29M18"
 33614 a 30508 c 27770 g 33925 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 60.5%; Score 24.2; DB 2; Length 125817;
 Matches 29; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY

2 atcaattgacctactctacctggctggctggactgggtg 38
 ||||| ||| ||||| ||||| ||||| |||||
 Db 83710 ATCACTTTACCCACCTCTACTTGGCTGAGCTGCGC 83674

RESULT

AC003111

LOCUS

AC003111 40649 bp DNA linear PRI 21-NOV-1997
 Human DNA from chromosome 19-specific cosmid R28194, genomic
 sequence, complete sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

HTG.
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 40649)
 Lamerdin, J.E., McCreedy, P.M., Adamson, A.W., Burkhardt-Schultz, K.,
 Gordon, L., Christensen, M., Kyle, A., Ramirez, M., Stilwagen, S.,
 Ganes, J., Danganan, L., Bruce, R., Quan, G., Montgomery, M., Ow, D.,
 Kobayashi, A., Olsen, A.O. and Carrano, A.V.

Sequence analysis of an OLFR cluster in 19p13

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (21-NOV-1997) Human Genome Center, Lawrence Livermore
 National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
 Map and sequence oriented from p telomere to centromere.
 Cosmid R28194 overlaps cosmid R31546 to the left and R28550 to the
 right.

FEATURES

source

1..40649
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="19"
 /map="19p13 from OLFR to D19S11"
 /clone="R28194"
 /cell_line="5HL2-B"
 /clone_lib="LL19NC03 R chromosome 19 specific cosmid
 library"
 /note="cosmid library constructed from LLNL from flow-sorted
 chromosomes from hybrid 5HL2-B, which carries chromosome
 19 as its only human chromosome."

repeat_region

repeat_region

repeat_region

repeat_region

misc_feature

complement(16..51)

/rpt_family="Alu"

complement(178..236)

/rpt_family="L1"

complement(263..542)

/rpt_family="Alu"

786..11357

/rpt_family="Alu"

complement(1885..2086)

/note="DPS similarity to sp|P25440|RING3_HUMAN RING3

PROTEIN (K1AA9001) gi|1082363|pir|A56619 female sterile

homeotic (fish) homolog RING3 - human; gi|31472 (X62083)

FSH gene product [Homo sapiens]; gi|

182769 (M80613) putative [Homo sapiens]; (67..1); 72%

identity.-(2211..1893) predicted exon, program:

grail2exons_human_1.3, frame: 2, quality: excellent.

score: 77.000-(2005..1893) DDS similarity to AA085917

2190809.r1 Stratagene colon (#937204) Homo sapiens CDNA

clone 511865 5' similar to gb:D42040 RING3 PROTEIN

(HUMAN);(1..113); 97% identity.-(1893..2050) DDS

similarity to AA087177 mk20h07.r1 Soares mouse p3NNF19.5

Mus musculus cDNA clone 493501 5' similar to gb:D42040

RING3 PROTEIN (HUMAN);(139..296); 80%

identity.-(2060..1893) DDS similarity to R22328 yh26e12.r1

Homo sapiens cDNA clone 130894 5' similar to gb:D42040

RING3 PROTEIN (HUMAN);(1..168); 100% identity."

complement(2216..2833)

/rpt_family="Alu"

complement(2972..3312)

/rpt_family="Alu"

complement(3438..3730)

/rpt_family="Alu"

4540..6221

/rpt_family="Tigger1"

4969..5271

/rpt_family="Alu"

complement(6137..6241)

/note="predicted exon, program: grail2exons_human_1.3,

frame: 1, quality: good, score: 52.000"

6241..6546

/rpt_family="Alu"

7699..7832

/note="predicted exon, program: grail2exons_human_1.3,

frame: 2, quality: good, score: 57.000"

complement(8546..8870)

/rpt_family="Alu"

complement(8885..9118)

/rpt_family="Alu"

complement(9393..10281)

/rpt_family="Alu"

complement(10842..11011)

/rpt_family="L1"

complement(10996..11281)

/rpt_family="Alu"

11281..11588

/note="DDS similarity to H00081 ph8c08u_19/1TV Homo

sapiens cDNA clone ph8c08u_19/1TV. Score: 642 Identity:

332/336 (98%).-(11404..114940) predicted exon, program:

grail2exons_human_1.3, frame: 0, quality: marginal, score:

45.000"

complement(11341..11582)

/rpt_family="Alu"

complement(12933..13215)

/rpt_family="Alu"

complement(15303..15511)

/rpt_family="Alu"

complement(15591..15874)

/rpt_family="Alu"

complement(18776..19074)

/rpt_family="Alu"

19211..19413

/note="predicted exon, program: grail2exons_human_1.3,

frame: 0, quality: good, score: 54.000"


```

402/404 (99%).
(1800..2152) R34240 yh80f09.s1 Homo sapiens cDNA clone
136073 3' similar to contains TAR1 repetitive element;
Score: 614 Identity: 346/363 (95%).
(2249..1897) IN71602 yw40b10.r1 Homo sapiens cDNA clone
254683 5' similar to contains element MER86 repetitive
element; Score: 588 Identity: 335/357 (93%).
(2212..1967) AA304114 EST16821 Aorta endothelial cells,
TNF alpha-treated Homo sapiens cDNA 5' end; Score: 443
Identity: 239/255 (93%).
1947..1973
/rpt_family="POLY_A"
2096..2176
/rpt_family="(CAAAA)n"
complement(2286..2339)
/rpt_family="POLY_A"
2313..3073
/Note="DSS similarity to overlapping ESTs:
(2340..2685) AA458860 aa26405.s1 NCI_CGAP_GCB1 Homo
sapiens cDNA clone IMAGE:814377 3'; Score: 683 Identity:
345/345 (100%).
(2787..2313) AA278877 zs79g12.r1 NCI_CGAP_GCB1 Homo
sapiens cDNA clone IMAGE:703750 5' similar to contains
element MER22 repetitive element; Score: 930 Identity:
473/475 (99%).
(2341..2692) AA573302 nm52e03.s1 NCI_CGAP_Br2 Homo sapiens
cDNA clone IMAGE:1071868; Score: 684 Identity: 347/352
(98%).
(2342..2769) AA579808 nk43f03.s1 NCI_CGAP_GC2 Homo sapiens
cDNA clone IMAGE:1016285; Score: 848 Identity: 426/428
(99%).
(2345..2777) AA121152 z188e01.s1 Stratagene colon
(#937204) Homo sapiens cDNA clone 511704 3'; Score: 842
Identity: 430/433 (99%).
(3003..2614) AA121132 z188e01.r1 Stratagene colon
(#937204) Homo sapiens cDNA clone 511704 5'; Score: 726
Identity: 383/386 (99%).
(2496..2807) T07432 EST05321 Homo sapiens cDNA clone
HFBE195; Score: 565 Identity: 306/317 (96%).
(2950..2560) R68397 yh99h03.r1 Homo sapiens cDNA clone
137909 5'. Score: 687 Identity: 382/396 (96%).
(3073..2821) T85539 yd78h03.r1 Homo sapiens cDNA clone
114389 5'. Score: 463 Identity: 244/254 (96%).
(3055..2902) AA666963 vr86d06.s1 Knowles Solter mouse 2
cell Mus musculus cDNA clone 1135595 5'; (355..514) 80%
identity.
And many others.
2510..2589
/rpt_family="POLY_A"
complement(3776..3858)
/rpt_family="(CAG)n"
complement(4402..4458)
/Note="DSS similarity to AA666963 vr86d06.s1 Knowles
Solter mouse 2 cell Mus musculus cDNA clone 1135595 5';
(298..354) 88% identity.
complement(4595..4757)
/Note="DSS similarity to AA666963 vr86d06.s1 Knowles
Solter mouse 2 cell Mus musculus cDNA clone 1135595 5';
(135..297) 86% identity.
complement(4846..4973)
/Note="DSS similarity to AA666963 vr86d06.s1 Knowles
Solter mouse 2 cell Mus musculus cDNA clone 1135595 5';
(1..134) 84% identity.
5268..5308
/rpt_family="MER5A"
complement(6747..7049)
/rpt_family="AluX"
8022..8145
/rpt_family="(TGG)n"
complement(8151..8405)
/Note="DSS similarity to AA057523 zf56e09.r1 Soares retina
N2b4HR Homo sapiens cDNA clone 380968 5' similar to
contains element MSRI repetitive element; Score: 434

```

```

repeat_region
9400..9540
/rpt_family="(TGG)n"
9540..9586
/rpt_family="(CAGA)n"
complement(10070..10174)
/rpt_family="MIR"
11994..13585

```

```

/Note="DSS similarity to overlapping ESTs:
(11994..12459) N49889 yv24b06.s1 Homo sapiens cDNA clone
243635 3'; Score: 893 Identity: 461/467 (98%).
(12541..12078) AA493873 nh03d10.s1 NCI_CGAP_Thy1 Homo
sapiens cDNA clone IMAGE:943219; Score: 809 Identity:
443/457 (96%).
(12197..12697) AA429349 zv50g01.s1 Soares ovary tumor
NbHOT Homo sapiens cDNA clone 757104 3'; Score: 978
Identity: 495/501 (98%).
(12197..12702) AA291401 zt38f02.s1 Soares ovary tumor
NbHOT Homo sapiens cDNA clone 724635 3'; Score: 976
Identity: 500/506 (98%).
(12197..12788) N45196 yz14a11.s1 Homo sapiens cDNA clone
283004 3'; Score: 1068 Identity: 573/584 (98%).
(12880..12348) AA625129 af70a09.r1 Soares NHMpu S1 Homo
sapiens cDNA clone 1047352 5'; Score: 1017 Identity:
527/528 (99%).
(13237..12768) AA291565 zt38f02.r1 Soares ovary tumor
NbHOT Homo sapiens cDNA clone 724635 5'; Score: 895
Identity: 466/469 (99%).
(13237..12850) AA428761 zv50g01.r1 Soares ovary tumor
NbHOT Homo sapiens cDNA clone 757104 5'; Score: 776
Identity: 388/388 (100%). -Other overlapping EST matches:
T51194"
complement(13455..>38069)
/gene="HUNK1"
/product="H. sapiens HUNK1 (Y12059)"
complement(13455..>38069)
/gene="HUNK1"
complement(13585..14213)
/gene="HUNK1"

```

mRNA

```

/gene="HUNK1"
/product="H. sapiens HUNK1 (Y12059)"
complement(13455..>38069)
/gene="HUNK1"
complement(13585..14213)
/gene="HUNK1"

```

misc_feature

```

/Note="DSS similarity to overlapping ESTs:
(13964..13585) R84483 y03ff05.r1 Homo sapiens cDNA clone
180033 5'; Score: 668 Identity: 360/380 (94%).
(14132..13748) AA349685 EST56553 Infant brain Homo sapiens
cDNA 5' end; Score: 743 Identity: 380/386 (98%).
(14213..13751) H30841 y078c10.r1 Homo sapiens cDNA clone
184050 5'; Score: 885 Identity: 457/462 (98%).
(14213..13839) R88182 ym90f03.r1 Homo sapiens cDNA clone
166205 5'; Score: 724 Identity: 372/377 (98%).
(14226..14171) R29034 F0-114D 22 week old human fetal
liver cDNA library Homo sapiens cDNA clone F0-114D 5';
(56..1) 100% identity.
And many others.
16375..16403
/rpt_family="POLY_G"
17570..18018
/Note="DSS similarity to AA262797 zs24f07.r1 NCI_CGAP_GCB1
Homo sapiens cDNA clone IMAGE:686149 5'; Score: 833
Identity: 435/446 (97%).
complement(18294..18623)

```

repeat_region

```

16375..16403
/rpt_family="POLY_G"
17570..18018

```

misc_feature

```

/Note="DSS similarity to AA262797 zs24f07.r1 NCI_CGAP_GCB1
Homo sapiens cDNA clone IMAGE:686149 5'; Score: 833
Identity: 435/446 (97%).
complement(18294..18623)

```

misc_feature

```

Query Match 59.5%; Score 23.8; DB 9; Length 42497;
Best Local Similarity 80.0%; Pred. No. 10;
Matches 28; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

```

Oy 3 tcacttgaccctactctacgtggctgactgggt 37
||||| ||| ||||| ||||| ||||| |||||

```

```

Db 36579 TCACGTGCACCAATCTCTAGCTGGCTGCACAGGT 36545

```

RESULT 4

```

AC109621/c AC109621
LOCUS AC109621 45420 bp DNA linear HTG 06-FEB-2002
DEFINITION Mus musculus clone RP23-330J17, LOW-PASS SEQUENCE SAMPLING.

```

AC109621
 VERSION
 HTG: HTGS_PHASE0.
 SOURCE
 house mouse.
 ORGANISM
 Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Mus musculus, clone RP23-330J17

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS

1 (bases 1 to 45420)
 2 (bases 1 to 45420)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhaltier,B.,
 Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
 Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A.,
 Cooke,P., DeArillano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
 Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Hagos,B., Horon,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R.,
 Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C.,
 MacDonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
 McKernan,P., McKernan,K., Meldrum,J., Meneus,L., Mihova,T.,
 Mlenga,Y., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
 Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
 Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
 Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
 Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
 Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zalnoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (06-FEB-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE
 JOURNAL
 COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: L20876
 Center clone name: 330_J_17

* NOTE: This record contains 57 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 753: contig of 753 bp in length
 * 754 853: gap of 100 bp
 * 854 1553: contig of 700 bp in length
 * 1554 1653: gap of 100 bp
 * 1654 2333: contig of 680 bp in length
 * 2334 2433: gap of 100 bp
 * 2434 3164: contig of 731 bp in length
 * 3165 3264: gap of 100 bp
 * 3265 3967: contig of 703 bp in length
 * 3968 4067: gap of 100 bp
 * 4068 4772: contig of 705 bp in length
 * 4773 4872: gap of 100 bp
 * 4873 5531: contig of 659 bp in length

5532 5631: gap of 100 bp
 * 5632 6291: contig of 660 bp in length
 * 6292 6391: gap of 100 bp
 * 6392 7098: contig of 707 bp in length
 * 7099 7198: gap of 100 bp
 * 7199 7883: contig of 685 bp in length
 * 7884 7983: gap of 100 bp
 * 7984 8697: contig of 714 bp in length
 * 8698 8797: gap of 100 bp
 * 8798 9506: contig of 709 bp in length
 * 9507 9606: gap of 100 bp
 * 9607 10317: contig of 711 bp in length
 * 10318 10417: gap of 100 bp
 * 10418 11109: contig of 692 bp in length
 * 11110 11209: gap of 100 bp
 * 11210 11918: contig of 709 bp in length
 * 11919 12018: gap of 100 bp
 * 12019 12709: contig of 691 bp in length
 * 12710 12809: gap of 100 bp
 * 12810 13509: contig of 700 bp in length
 * 13510 13609: gap of 100 bp
 * 13610 14299: contig of 690 bp in length
 * 14300 14399: gap of 100 bp
 * 14400 15105: contig of 706 bp in length
 * 15106 15205: gap of 100 bp
 * 15206 15908: contig of 703 bp in length
 * 15909 16008: gap of 100 bp
 * 16009 16713: contig of 705 bp in length
 * 16714 16813: gap of 100 bp
 * 16814 17522: contig of 709 bp in length
 * 17523 17622: gap of 100 bp
 * 17623 18299: contig of 677 bp in length
 * 18300 18399: gap of 100 bp
 * 18400 19099: contig of 700 bp in length
 * 19100 19199: gap of 100 bp
 * 19200 19903: contig of 704 bp in length
 * 19904 20003: gap of 100 bp
 * 20004 20699: contig of 696 bp in length
 * 20700 20799: gap of 100 bp
 * 20800 21508: contig of 709 bp in length
 * 21509 21608: gap of 100 bp
 * 21609 22316: contig of 708 bp in length
 * 22317 22416: gap of 100 bp
 * 22417 23121: contig of 705 bp in length
 * 23122 23221: gap of 100 bp
 * 23222 23888: contig of 667 bp in length
 * 23889 23988: gap of 100 bp
 * 23989 24689: contig of 701 bp in length
 * 24690 24789: gap of 100 bp
 * 24790 25491: contig of 702 bp in length
 * 25492 25591: gap of 100 bp
 * 25592 26254: contig of 663 bp in length
 * 26255 26354: gap of 100 bp
 * 26355 27058: contig of 704 bp in length
 * 27059 27158: gap of 100 bp
 * 27159 27861: contig of 703 bp in length
 * 27862 27961: gap of 100 bp
 * 27962 28666: contig of 705 bp in length
 * 28667 28766: gap of 100 bp
 * 28767 29476: contig of 710 bp in length
 * 29477 29576: gap of 100 bp
 * 29577 30272: contig of 696 bp in length
 * 30273 30372: gap of 100 bp
 * 30373 31062: contig of 690 bp in length
 * 31063 31162: gap of 100 bp
 * 31163 31867: contig of 705 bp in length
 * 31868 31967: gap of 100 bp
 * 31968 32670: contig of 703 bp in length
 * 32671 32770: gap of 100 bp
 * 32771 33475: contig of 705 bp in length
 * 33476 33575: gap of 100 bp
 * 33576 34289: contig of 714 bp in length
 * 34290 34389: gap of 100 bp

```
* 34390 35109: contig of 720 bp in length
* 35110 35209: gap of 100 bp
* 35210 35908: contig of 699 bp in length
* 35909 36008: gap of 100 bp
* 36009 36689: contig of 681 bp in length
* 36690 36789: gap of 100 bp
* 36790 37484: contig of 695 bp in length
* 37485 38238: contig of 654 bp in length
* 38239 38338: gap of 100 bp
* 38339 39006: contig of 668 bp in length
* 39007 39106: gap of 100 bp
* 39107 39805: contig of 699 bp in length
* 39806 39905: gap of 100 bp
* 39906 40618: contig of 713 bp in length
* 40619 40718: gap of 100 bp
* 40719 41411: contig of 693 bp in length
* 41412 41511: gap of 100 bp
* 41512 42211: contig of 700 bp in length
* 42212 42311: gap of 100 bp
* 42312 43010: contig of 699 bp in length
* 43011 43110: gap of 100 bp
* 43111 43823: contig of 713 bp in length
* 43824 43923: gap of 100 bp
* 43924 44627: contig of 704 bp in length
* 44628 44727: gap of 100 bp
* 44728 45420: contig of 693 bp in length.
```

```
FEATURES
    source
    1. 45420
    /organism="Mus musculus"
    /db_xref="taxon:10090"
    /clone="RP23-330J17"
    /clone_lib="RPCI-23 Female Mouse BAC"
BASE COUNT 11013 a 8360 c 8491 g 11360 t 6196 others
ORIGIN
```

```
Query Match 57.0%; Score 22.8; DB 2: Length 45420;
Best Local Similarity 79.4%; Pred. NO. 28;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
```

```
Qy 4 cacttgaccctactactctggctggctgactgggt 37
| | | | | | | | | | | | | | | | | | | | |
```

```
Db 18473 CCCATGACCCCTACTCTCTAGGTAGGACTGGAT 18440
```

```
RESULT 5
AC019031 AC019031 157076 bp DNA linear PRI 01-DEC-2001
LOCUS Homo sapiens chromosome 8, clone RP11-563019, complete sequence.
DEFINITION AC019031
ACCESSION AC019031
VERSION AC019031.7 GI:16974237
KEYWORDS HTG.
SOURCE human.
```

```
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 157076)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferrelira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Hearford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
```

```
2 (bases 1 to 157076)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferrelira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Hearford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
```

Norman,C.H., O'Connor,T., O'Donnell,P., Oliviar,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.

Direct Submission

Submitted (30-DEC-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE

AUTHORS

3 (bases 1 to 157076)
Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferrelira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gird,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Hearford,A., Horton,L., Hulme,W., Iliev,R., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazaras,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
McCarthy,M., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,F., McPheeters,R., Meldrim,J.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Olivier,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triglilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zaloun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (01-DEC-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

On Nov 18, 2001 this sequence version replaced gi:16041469.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L4200

Center clone name: 563_O_19

FEATURES

source

Location/Qualifiers

1. 157076

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="8"

/map="8"

/clone="RP11-563019"

/clone_lib="RPCI-11 Human Male BAC"

8. 2498

/rpt_family="L1MB4"

2532. 2698

/rpt_family="L1MB4"

3183. 3281

/rpt_family="MLTIG"

3288. 3810

/rpt_family="MLTIF1"

3846. 4155

/rpt_family="MLTIG"

7069. 7096

/rpt_family="AT-rich"

complement(8212..8404)

/rpt_family="MIR3"

9488. 9617

/rpt_family="MER5A"

9849. 9869

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region


```

ORGANISM      Homo sapiens
REFERENCE     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS       Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE         1 (bases 1 to 213611)
JOURNAL       Birren,B., Linton,L., Nusbaum,C. and Lander,E.
REFERENCE     Homo sapiens chromosome 8, clone RP11-420F14
AUTHORS       Unpublished
TITLE         2 (bases 1 to 213611)
JOURNAL       Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
AUTHORS       Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B., Brown,A.,
REFERENCE     Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B., Brown,A.,
AUTHORS       Canarata,J., Campolano,A., Chang,J., Choepel,Y., Colangelo,M.,
TITLE         Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K.,
JOURNAL       Diaz,J.S., Dodge,S., Fero,S., Ferreira,P., FitzHugh,W., Gage,D.,
AUTHORS       Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
REFERENCE     Grand-Pierre,N., Hagos,B., Heaford,A., Horton,D., Hulme,W.,
AUTHORS       Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
TITLE         Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
JOURNAL       McClean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
AUTHORS       Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
REFERENCE     Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
AUTHORS       Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
TITLE         Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
JOURNAL       Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S.,
AUTHORS       Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
REFERENCE     Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
AUTHORS       Theodore,J., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,
TITLE         Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
JOURNAL       Zalnoun,J., Zembek,L., Zimmer,A. and Zody,M.
REFERENCE     Direct Submission
AUTHORS       Submitted (07-MAY-2001) Whitehead Institute/MIT Center for Genome
JOURNAL       Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT       On Feb 1, 2002 this sequence version replaced gi:17975289.
              All repeats were identified using RepeatMasker:
              Smit, A.F.A. & Green, P. (1996-1997)
              http://ftp.genome.washington.edu/RM/RepeatMasker.html
              ----- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIDR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
              ----- Project Information
              Center project name: L12187
              Center clone name: 420_F_14
              -----
              * NOTE: This is a 'working draft' sequence. It currently
              * consists of 3 contigs. The true order of the pieces
              * is not known and their order in this sequence record is
              * arbitrary. Gaps between the contigs are represented as
              * runs of N, but the exact sizes of the gaps are unknown.
              * This record will be updated with the finished sequence
              * as soon as it is available and the accession number will
              * be preserved.
              *
              * 1 49517: contig of 49517 bp in length
              * 49518 49617: gap of 100 bp
              * 49618 143781: contig of 94164 bp in length
              * 143782 143881: gap of 100 bp
              * 143882 213611: contig of 69730 bp in length.
              Location/Qualifiers
                1..213611
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /chromosome="8"
                  /map="8"
                  /clone="RP11-420F14"
                  /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    57973 a 50973 c 49786 g 54679 t 200 others
ORIGIN

```

```

Query Match      57.0%; Score 22.8; DB 2; Length 213611;
Best Local Similarity 79.4%; Pred No. 28;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

```

QY      6 cttgaccctactctactctgggtgactggactgggtgg 39
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 198343 CTTGGCCCTACTCTGCTTGGCAGGAGTGAGGG 198310

RESULT      8
AF276759    AF276759      213621 bp      DNA      linear      PRI 06-FEB-2002
LOCUS      Homo sapiens chromosome 8 clone RP11-420F14 map 8p12, complete
DEFINITION sequence.
ACCESSION  AF276759
VERSION    AF276759.2 GI:18543386
KEYWORDS  HTG: HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 213621)
AUTHORS   Wen,G., Reichwald,K., Baumgart,C., Taudin,S., Baumgart,C. and
           Platzer,M.
TITLE     Chromosome 8 genomic sequence
JOURNAL   Unpublished
AUTHORS   2 (bases 1 to 213621)
JOURNAL   Genome Sequencing Center Jena.
AUTHORS   Direct Submission
JOURNAL   Submitted (09-JUN-2000) Genome Analysis, Institute of Molecular
           Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
REFERENCE  3 (bases 1 to 213621)
AUTHORS   Wen,G. and Platzer,M.
JOURNAL   Direct Submission
COMMENT    Submitted (06-FEB-2002) Genome Analysis, Institute of Molecular
           Biotechnology, Beutenbergstr. 11, Jena, Thuringia 07745, Germany
           On Feb 6, 2002 this sequence version replaced gi:8572764.
           ----- Genome Center
           Center: Institute of Molecular Biotechnology
           Center code: IMB
           Web site: http://genome.imb-jena.de/
           Contact: gscj-submit@genome.imb-jena.de
           ----- Project Information
           Center project name: H534
           Center clone name: RP11-420F14
           ----- Summary Statistics
           Sequencing vector: M13; 100% of reads
           Chemistry: Dye-terminator Big Dye; 100% of reads
           Assembly program: Phrap; version 0.990329
           Consensus quality: 200655 bases at least Q40
           Consensus quality: 205435 bases at least Q30
           Consensus quality: 208300 bases at least Q20
           Quality coverage: 4.63 x in Q20 bases; sum-of-contigs
           -----
           Sequence Quality Assessment:
           This entry has been annotated with sequence quality
           estimates computed by the Phrap assembly program.
           All manually edited bases have been reduced to quality 10.
           Quality levels above 40 are expected to have less than
           1 error in 10,000 bp.
           Base-by-base quality values are not generally visible from the
           GenBank flat file format but are available as part
           of this entry's ASN.1 file.
           -----
           Location/Qualifiers
             1..213621
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /chromosome="8"
               /map="8p12"
               /clone="RP11-420F14"
BASE COUNT    54747 a 49879 c 50913 g 58082 t
ORIGIN

```

```

Query Match      57.0%; Score 22.8; DB 9; Length 213621;

```


misc_feature	1517. .1610 /note="Sequence from overlapping clone AL358092. Assembly confirmed by restriction digest"
repeat_region	1739. .1840 /note="MER91C repeat: matches 1. .108 of consensus"
repeat_region	3406. .3717 /note="AluY repeat: matches 1. .311 of consensus"
repeat_region	3723. .3746 /note="6 copies 4 mer acac 100% conserved"
repeat_region	3996. .4360 /note="MLT1E repeat: matches 183. .568 of consensus"
repeat_region	4737. .4882 /note="MIR repeat: matches 19. .182 of consensus"
repeat_region	4955. .5460 /note="LTR33 repeat: matches 1. .531 of consensus"
repeat_region	7652. .7974 /note="MLT1I repeat: matches 43. .403 of consensus"
repeat_region	10608. .10748 /note="MERSA repeat: matches 12. .171 of consensus"
misc_feature	10708. .11070 /note="match: GSS: Em:AQ045804"
repeat_region	11233. .11530 /note="Alusg repeat: matches 1. .289 of consensus"
repeat_region	11806. .11852 /note="L2 repeat: matches 2699. .2749 of consensus"
repeat_region	12672. .12701 /note="10 copies 3 mer att 90% conserved"
misc_feature	13520. .14027 /note="match: GSS: Em:AQ684879"
misc_feature	13530. .13960 /note="match: GSS: Em:AQ142591"
misc_feature	13533. .14033 /note="match: GSS: Em:AQ358097"
misc_feature	13568. .14117 /note="match: GSS: Em:AQ178917"
repeat_region	13828. .14100 /note="AluY repeat: matches 39. .310 of consensus"
repeat_region	14956. .15018 /note="LTR33 repeat: matches 423. .479 of consensus"
repeat_region	15019. .15411 /note="MSTA repeat: matches 1. .425 of consensus"
repeat_region	15412. .15450 /note="LTR33 repeat: matches 479. .519 of consensus"
repeat_region	15763. .16395 /note="L1ME2 repeat: matches 5485. .6155 of consensus"
repeat_region	17415. .17756 /note="L2 repeat: matches 2352. .2735 of consensus"
repeat_region	18016. .18323 /note="AluSx repeat: matches 1. .312 of consensus"
repeat_region	20169. .20381 /note="L1MB3 repeat: matches 5967. .6183 of consensus"
repeat_region	20382. .20452 /note="L1MB3 repeat: matches 5967. .6038 of consensus"
repeat_region	20453. .21033 /note="L1MB3 repeat: matches 5431. .6038 of consensus"
repeat_region	21062. .21377 /note="AluY repeat: matches 1. .311 of consensus"
repeat_region	22624. .22750 /note="MER45 repeat: matches 50. .178 of consensus"
repeat_region	23371. .24290 /note="46 copies 20 mer 83% conserved"
repeat_region	24399. .24639 /note="L2 repeat: matches 2251. .2495 of consensus"
repeat_region	25098. .25157 /note="30 copies 2 mer ac 90% conserved"
repeat_region	25165. .25500 /note="MER2 repeat: matches 4. .343 of consensus"
repeat_region	25539. .25631 /note="MIR repeat: matches 106. .201 of consensus"
misc_feature	26339. .27004 /note="match: GSS: Em:AQ473715"
repeat_region	26480. .26507 /note="14 copies 2 mer ca 96% conserved"


```

/note="Genefinder prediction"
1984..2091
/note="Genefinder prediction"
complement(2020..2135)
/note="XPOUND prediction, score = 0.379"
2488..2669
/note="homology = 79.1%, counts = 13"
/rpt_family="atacacacacac repeat"
/rpt_type=TANDEM
2506..2600
/note="AC repeat"
2530..2582
/note="IR1, 79% complementary to IR1' (3707..3759)"
/rpt_type=INVERTED
3227..3243
/note="AT repeat"
complement(3248..3302)
/note="match: GSS AQ213585"
3256..3302
/note="match: GSS AQ260331"
complement(3435..3561)
/note="84% identity: matches 134..260 of consensus"
/rpt_family="AluJb"
3440..3537
/note="Genefinder prediction"
complement(3613..3672)
/note="match: GSS AQ234268"
3694..3833
/note="homology = 60.0%, counts = 70"
/rpt_family="at repeat"
/rpt_type=TANDEM
3707..3759
/note="IR1', 79% complementary to IR1 (2530..2582)"
/rpt_type=INVERTED
4075..4522
/note="match: EST A1859260"
4101..4137
/note="WZEF prediction, score = 0.907"
complement(4698..4711)
/note="XPOUND prediction, score = 0.816"
complement(5780..5943)
/note="WZEF prediction, score = 0.811"
complement(5796..5946)
/note="86% identity: matches 245..395 of consensus"
/rpt_family="LI"
5796..5946
/note="90% identity: matches 130..280 of consensus"
/rpt_family="AluSb"
complement(5822..5919)
/note="Genefinder prediction"
5950..6001
/rpt_family="homology = 84.6%, counts = 4"
/rpt_type=TANDEM
6683..6695
/note="TC repeat"
complement((7117,7208)..7546)
/note="match: GSS B14745 B59625"
complement(7586..7640)
/note="WZEF prediction, score = 0.568"
7791..7803
/note="TC repeat"
complement(8072..8088)
/note="XPOUND prediction, score = 0.241"
8090..8103
/note="TC repeat"
8178..8217
/note="TTTA repeat"
8178..8225
/note="homology = 95.8%, counts = 12"
/rpt_family="titia repeat"
/rpt_type=TANDEM
8267..8450

```

```

repeat_region
/note="CpG_island (%GC=59.8, o/e=1.07, #CpGs=15)"
complement(8268..8364)
/note="97% identity: matches 132..228 of consensus"
/rpt_family="AluSb"
8273..8364
/note="91% identity: matches 302..393 of consensus"
/rpt_family="LI"
complement(8274..8463)
/note="Genefinder prediction"
complement(8381..8497)
/note="95% identity: matches 1..117 of consensus"
/rpt_family="AluSb"
8381..8486
/note="95% identity: matches 410..515 of consensus"
/rpt_family="LI"
complement(8399..8435)
/note="XPOUND prediction, score = 0.347"
complement(8860..8991)
/note="GRAIL, score = 97%, comment = excellent"
complement(8873..8934)
/note="XPOUND prediction, score = 0.391"
complement(8874..8947)
/note="90% identity: matches 310..383 of consensus"
/rpt_family="LI"
8972..9037
/note="87% identity: matches 3445..3510 of consensus"
/rpt_family="LI"
complement((9043,9056)..(9084,9089))
/note="match: GSS AQ394050 AQ114719 AQ505132 AQ745232
AQ417781 AQ630059 AQ100697 B75664 AQ486114 AQ531167"
complement(9044..9086)
/note="match: EST A1810082"
(9045,9047)..(9086,9089)
/note="match: GSS AQ538467 B75590 AQ584427 AQ348442

Query Match      56.5%; Score 22.6; DB 9; Length 178660;
Best Local Similarity 75.7%; Pred. No. 35;
Matches 28; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 3 tcacttgaccctactctacctggctggactgggtg9 39
||| | ||| ||||| ||||| |||||
Db 125893 TCCTTAAACTAGTCTACCTAGGCTGGGCTGGTGG 125929

RESULT 14
AP001700      340000 bp      DNA      linear      PRI 30-MAY-2000
LOCUS      Homo sapiens genomic DNA, chromosome 21q, section 44/105.
DEFINITION      AP001700 AL163245 BA000005
ACCESSION      AP001700.1 GI:7768710
VERSION
KEYWORDS
SOURCE      Homo sapiens DNA.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS      Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., and Schudy,A.
The DNA sequence of human chromosome 21. The chromosome 21 mapping
and sequencing consortium
Nature 405 (6784), 311-319 (2000)
20289799
2 (bases 1 to 340000)
Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S.,

```

Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordstiek,G., Hornischer,K., Barandt,P., Scharfe,M., Schoen,O., Desario,A., Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S., Hennig,S., Riesenmann,L., Dagand,E., Wehrmaeyer,S., Borzym,K., Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and Vaspo,M.L.

TITLE JOURNAL

Direct Submission
Submitted (10-APR-2000) to the DDBJ/EMBL/GenBank databases. The Chromosome 21 Mapping and Sequencing Consortium: * RIKEN Genomic Sciences Center, Human Genome Research Group * Institute of Molecular Biotechnology, Genome Analysis * Keio University School of Medicine, Dept. of Molecular Biology * GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular Genetics (addresses see below)

COMMENT

On May 30, 2000 this sequence version replaced gi:7717301.
The chromosome 21 mapping and sequencing consortium consisting of
* RIKEN Genomic Sciences Center, Human Genome Research Group, * Sagamihara 228-8555, Japan,
* e.mail: hattorie@gsr.riken.go.jp
* URL: http://hgp.gsc.riken.go.jp/
and
* Institute of Molecular Biotechnology, Genome Analysis, * Beutenbergstrasse 11, D-07745 Jena, Germany,
* e.mail: gscj-submit@genome.imb-jena.de
* URL: http://genome.imb-jena.de/
and
* Keio University School of Medicine, Molecular Biology, * Tokyo 160-8502, Japan,
* e.mail: nshimizudmb-med.keio.ac.jp
* URL: http://www.dmb.med.keio.ac.jp/
and
* GBF, Dept. of Genome Analysis,
* Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail: info.genome@gbf.de
* URL: http://genome.gbf.de/
and
* Max-Planck Institute for Molecular Genetics,
* Ihnestrasse 73, D-14195 Berlin, Germany,
* e.mail: info-chr21@molgen.mpg.de
* URL: http://chr21.rz-berlin.mpg.de/
AL163245: Submitted (10-Apr-2000).
Location/Qualifiers
1..340000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
<1..104356
/note="Accession No. AP001607"
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1"
/clone="KB851D4, 5' partial"
/clone_lib="Keio BAC library"
263..627
/note="(TA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
920..1710
/note="L1PB1"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
1720..1752
/note="(TG)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
1753..1834
/note="(CA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
1835..2193
/note="L1PB1"
/rpt_family="LINE/L1"

FEATURES

source

source

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region
/rpt_type=DISPERSED
2209..2521
/note="L1p"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
2535..2818
/note="AluY"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(2819..3122)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
complement(3988..4066)
/note="MER115"
/rpt_family="DNA"
/rpt_type=DISPERSED
complement(4100..4253)
/note="MER5b"
/rpt_family="DNA/MER1_type"
/rpt_type=DISPERSED
5543..6453
/note="L1MA5A"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
6798..6847
/note="L1MA4A"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
complement(6954..7127)
/note="MER20"
/rpt_family="DNA/MER1_type"
/rpt_type=DISPERSED
7730..7957
/note="HAL1"
/rpt_family="LINE/Other"
/rpt_type=DISPERSED
8282..8343
/note="L1PA3"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
complement(8503..8813)
/note="AluY"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(8875..9276)
/note="L1ME1"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
9581..9944
/note="MLT1-INTERNAL"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
10088..10623
/note="MLT1-INTERNAL"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
10699..11123
/note="MLTID"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
11589..11659
/note="MLTID"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
11661..11847
/note="HAL1"
/rpt_family="LINE/Other"
/rpt_type=DISPERSED
12059..12101
/note="(CAA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM

```
repeat_region 12144..12172
/note="(CA)n"
/rpt_family="Simple_repeat"
repeat_region /rpt_type=TANDEM
complement(12195..12258)
/note="MIR"
/rpt_family="SINE/MIR"
repeat_region /rpt_type=DISPERSED
13911..14148
/note="L2"
/rpt_family="LINE/L2"
repeat_region /rpt_type=DISPERSED
14541..14917
/note="L2"
/rpt_family="LINE/L2"
repeat_region /rpt_type=DISPERSED
complement(14918..15219)
/note="AluSp"
/rpt_family="SINE/Alu"
repeat_region /rpt_type=DISPERSED
15220..15246
/note="L2"
/rpt_family="LINE/L2"
repeat_region /rpt_type=DISPERSED
16873..17238
/note="THE1C"
/rpt_family="LTR/MaLR"
```

Query Match 56.5%; Score 22.6; DB 9; Length 340000;
Best Local Similarity 75.7%; Pred. No. 35;
Matches 28; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```
QY 3 tcaattgacctactctacctggctggactgggtgg 39
||||| ||| ||||| ||||| ||||| |||||
Db 310955 TCCTTAAACTAGTCTACCTAGGCTGGCGGTGGTGG 310991
```

RESULT 15

G42160
LOCUS G42160 333 bp DNA linear STS 29-OCT-1998
DEFINITION ps701068 KWOK Homo sapiens STS genomic, sequence tagged site.
ACCESSION G42160
VERSION G42160.1 GI:3808205
KEYWORDS STS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 333)
Zakeri,H., Gu,Z., Hillier,L. and Kwok,P.-Y.
TITLE Single nucleotide polymorphisms from genomic sequences
JOURNAL Unpublished
COMMENT

Contact: Hamideh Zakeri
Pui-Yan Kwok
Washington University School of Medicine
Dermatology, Box 8123, 660 South Euclid Avenue, St. Louis, MO
63110, USA
Tel: 314-362-8162
Fax: 314-362-8159
Email: hamideh@psts.wustl.edu
Primer A: TGACATGATGACCAAGCACC
Primer B: GGAAAGCTGTCTCAGCATCCAG
STS size: 333
PCR profile:
Presoak: 93 degrees C for 1.00 minute
Denaturation: 93 degrees C for 10 seconds
Annealing: 60 degrees C for 1.00 minute
Polymerization: 72 degrees C for 1.00 minute
Final Extension: 72 degrees C for 5.00 minutes
PCR cycles: 35
Thermal Cycler: Perkin-Elmer 9600
Protocol:

Template: 16 ng Human genomic DNA
Primer: each 0.8 uM
dNTPs: each 200 uM
Amplitaq Polymerase: 2 units
Total Vol: 40 ul

Buffer:
MgCl2: 1.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 8.3.

FEATURES

Source 1..333 Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="KWOK"
STS 1..333
primer_bind 1..22
primer_bind complement(312..333) 81 t
BASE COUNT 86 a 74 c 92 g 81 t
ORIGIN

Query Match 56.0%; Score 22.4; DB 11; Length 333;
Best Local Similarity 81.2%; Pred. No. 40;
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 5 actgagccctactctacctggctggactggg 36
||||| ||||| ||| ||||| ||||| |||
Db 132 ACTGTGCCCTTCACCTGGGCTGGAGTGTG 163

Search completed: June 4, 2002, 17:47:45
Job time: 6733 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 17:08:54 ; Search time 4272.65 Seconds
(without alignments)
126.357 Million cell updates/sec

Title: US-09-721-543A-13

Perfect score: 40
Sequence: 1 ctagagccaccataaccctcaataactccaggattggg 40

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estinu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_htc:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_htc:*
- 12: gb_gss:*
- 13: em_gss_hum:*
- 14: em_gss_inv:*
- 15: em_gss_pln:*
- 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	24.6	61.5	296	9 AW845395	AW845395 CM4-CT004
C 2	24.6	61.5	486	10 BF740572	BF740572 QV1-HB003
C 3	23.8	59.5	250	10 T08487	T08487 EST06378 In
C 4	23.6	59.0	337	10 T92497	T92497 ye24b09.r1
C 5	23	57.5	69	10 BE878784	BE878784 601493062
C 6	23	57.5	118	9 AW804230	AW804230 PM3-UM008
C 7	23	57.5	135	9 BE005490	BE005490 CM1-BN011
C 8	23	57.5	147	9 BE005492	BE005492 CM1-BN011
C 9	23	57.5	161	10 N73434	N73434 yz3le12.r1
C 10	23	57.5	208	10 BF090424	BF090424 RCI-NT003
C 11	23	57.5	223	9 BE089733	BE089733 QV0-BT070
C 12	23	57.5	225	10 BF340702	BF340702 602035458
C 13	23	57.5	227	9 BE005486	BE005486 CM1-BN011
C 14	23	57.5	231	9 AA327421	AA327421 EST30756
C 15	23	57.5	237	10 BF037734	BF037734 601462081
C 16	23	57.5	243	10 BG105201	BG105201 602312948
C 17	23	57.5	249	9 AW804233	AW804233 PM3-UM008

C 18	57.5	251	10 BE622136	BE622136 601440862
C 19	57.5	252	10 BG288188	BG288188 602383704
C 20	57.5	252	10 BE869246	BE869246 601445061
C 21	57.5	252	10 BE894967	BE894967 601436172
C 22	57.5	260	9 AA101534	AA101534 zn79c11.s
C 23	57.5	263	9 AA383010	AA383010 EST96488
C 24	57.5	263	10 BE878754	BE878754 601493017
C 25	57.5	273	10 BF848202	BF848202 QV0-EN005
C 26	57.5	276	10 BE617770	BE617770 601441968
C 27	57.5	277	10 BG290247	BG290247 602384948
C 28	57.5	289	10 T39447	T39447 ya06a08.r1
C 29	57.5	290	10 BF740849	BF740849 QV1-HB003
C 30	57.5	299	9 BE178852	BE178852 PM1-HT060
C 31	57.5	304	9 AA384401	AA384401 EST97914
C 32	57.5	305	9 AA853875	AA853875 NHTBCae09
C 33	57.5	307	10 AW804241	AW804241 PM3-UM008
C 34	57.5	307	10 T39458	T39458 ya06b07.r2
C 35	57.5	308	9 AW804238	AW804238 PM3-UM008
C 36	57.5	308	9 AA383479	AA383479 EST96859
C 37	57.5	313	10 BE936437	BE936437 RCI-NT003
C 38	57.5	316	10 BF342651	BF342651 602013792
C 39	57.5	317	10 BG169657	BG169657 602324628
C 40	57.5	318	10 BF968657	BF968657 602270903
C 41	57.5	321	10 BI015712	BI015712 PM0-ET020
C 42	57.5	322	10 BI013197	BI013197 PM0-ET020
C 43	57.5	322	10 BI013200	BI013200 PM0-ET020
C 44	57.5	324	10 BF884828	BF884828 PM0-ET020
C 45	57.5	326	9 AA361993	AA361993 EST71440

ALIGNMENTS

RESULT 1

AW845395/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES

source

/organism="Homo sapiens"

AW845395 295 bp mRNA linear EST 19-MAY-2000
CM4-CT0045-180200-512-a09 CT0045 Homo sapiens cDNA, mRNA sequence.
AW845395
AW845395.1 GI:7940912
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 295)
Dias Neto E., Garcia Correa R., Verjovski-Almeida S., Briones M.R.,
Nagai M.A., da Silva W. Jr., Zago M.A., Bordin S., Costa F.F.,
Goldman G.H., Carvalho A.F., Matsukuma A., Balg G.S., Simpson D.H.,
Brunstein A., de Oliveira P.S., Bucher P., Jongeneel C.V., O'Hare
M.J., Soares F., Brentani R.R., Reis L.F., de Souza S.J. and
Simpson A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2=CM4-CT0045-180
200-512-a09&t3=2000-02-18&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 46
High quality sequence stop: 200.
Location/Qualifiers
i .296
/organism="Homo sapiens"

```

BASE COUNT      129 a   129 c   110 g   118 t
ORIGIN

Query Match          61.5%; Score 24.6; DB 10; Length 486;
Best Local Similarity 76.9%; Pred. No. 35;
Matches 30; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY      2 tcgagccaccacccataaacctcctaactccaggagattggg 40
db      167 TCCAGCCACACTTGGAAGCCACAAAACCTGCAGGAGATTGGG 205

```

RESULT	3
LOCUS	T08487
DEFINITION	EST06378 Infant Brain, Bento Soares Homo sapiens cDNA clone HIBBE96 5' end, mRNA sequence.
ACCESSION	T08487
VERSION	T08487.1
KEYWORDS	GI:389515
SOURCE	EST. human.
	250 bp mRNA linear EST 03-AUG-1993

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 250)
ADAMS, N.D., SOARES, M.B., KARLAVAGE, A.R., FIELDS, C. and VENTER, J.C.
Rapid cDNA sequencing (expressed sequence tags) from a
directionally cloned human infant brain cDNA library
Nature Genet. 4, 373-380 (1993)
JOURNAL
MEDLINE
94004965

CONTACT	<p>Contact: Adams, MD The Institute for Genomic Research 932 Clopper Road, Gaithersburg, MD 20878 Tel: 3018599056 Fax: 3018599423 Email: mdadams@tigr.org Seq primer: M13 Reverse.</p>
FEATURES	<p>Location/Qualifiers 1..250</p>
source	

BASE COUNT		ORIGIN	
56 a	84 c	72 g	31 t
		7 others	
Query Match 59.5%; Score 23.8; DB 10; Length 250; Best Local Similarity 77.8%; Pred. No. 60; Matches 28; Conservative 0; Mismatches 8; Indels 0; Gaps 0;			
Qy		4 gagccaccacataaacctcaataactccaggattgg 39	

RESULT	4
T92497/c	
LOCUS	
DEFINITION	
IMAGE: y24b03.rl Stragatene lung (#37210) Homo sapiens cDNA clone	
IMAGE: I18649 5', mRNA sequence.	
ACCESSION	T92497
VERSION	T92497.1 GI:724410
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 337)
AUTHORS	Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins


```
aps 0;
```

-OCT-2000
69 5'

MGC)

NOTES;
1100 DT.

aps
0:

aps
0:

THIS PAGE BLANK (USPTO)

Result No.	Query			DB	ID	Description	
	Score	Match	Length				
C 1	23	57.5	2634	2	US-08-818-514-1	Sequence 1, Appli	
	2	57.5	2634	2	US-08-818-514-2	Sequence 2, Appli	
	3	23	57.5	2634	3	US-09-115-934A-1	Sequence 1, Appli
C 3	4	23	57.5	2634	3	US-09-115-934A-2	Sequence 2, Appli
	5	20	50.0	3364	2	US-08-735-609-9	Sequence 9, Appli
	6	20	50.0	3364	2	US-08-735-609-9	Sequence 9, Appli
C 7	20	50.0	3364	2	US-09-315-372-9	Sequence 9, Appli	
	8	20	50.0	3364	3	US-09-244-752-9	Sequence 9, Appli
	9	20	50.0	3364	3	US-09-244-752-9	Sequence 9, Appli
C 10	10	19.2	48.0	1864	3	US-09-221-235-10	Sequence 10, Appli
	11	19.2	48.0	1864	3	US-09-221-928-10	Sequence 10, Appli
	12	19.2	48.0	1864	3	US-09-221-527-10	Sequence 10, Appli
C 13	13	19.2	48.0	1864	3	US-09-221-236-10	Sequence 10, Appli
	14	19.2	48.0	1864	3	US-09-221-416-10	Sequence 10, Appli
	15	19.2	48.0	1864	4	US-09-221-245-10	Sequence 10, Appli
C 16	16	19.2	48.0	1864	4	US-09-163-115-10	Sequence 10, Appli
	17	19.2	48.0	1864	4	US-09-221-528-10	Sequence 10, Appli
	18	19.2	48.0	1864	4	US-09-593-553-10	Sequence 10, Appli
C 19	19	19.2	48.0	1864	4	US-09-221-373-10	Sequence 10, Appli
	20	18.8	47.0	32207	2	US-08-770-379-20	Sequence 20, Appli
	21	18.8	47.0	32207	4	US-08-757-669A-20	Sequence 20, Appli
C 22	22	18.8	47.0	32207	4	US-09-230-371A-20	Sequence 20, Appli
	23	18.6	46.5	798	2	US-08-742-009-2	Sequence 2, Appli
	24	18.6	46.5	798	3	US-09-172-110-2	Sequence 2, Appli
C 25	25	18.4	46.0	1045	1	US-08-317-880-3	Sequence 3, Appli
	26	18.4	46.0	1045	2	US-08-782-396-3	Sequence 3, Appli
	27	18.4	46.0	1346	4	US-09-364-230-5	Sequence 5, Appli

```
RESULT 2
US-08-818-514-2
; Sequence 2, Application US/08818514
; Patent No. 5837838
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; TITLE OF INVENTION: BAX Inhibitor Proteins
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,514
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2446
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2634 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-818-514-2

Query Match 57.5%; Score 23; DB 2; Length 2634;
Best Local Similarity 74.4%; Pred. No. 0.81;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 tcgagccaccataaccctcaactccaggattggg 40
||| ||||| || ||||| ||||| ||||| |||||
Db 579 TCCACCCACACTTGAAGCCACAAAACTCAGGATTGGG 617

RESULT 3
US-09-115-934A-1/c
; Sequence 1, Application US/09115934A
; Patent No. 6130317
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Xu, Qunli
; TITLE OF INVENTION: BAX Inhibitor Proteins
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/115,934A
```

```
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/818,514
; FILING DATE: 14-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3209
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2634 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-115-934A-1

Query Match 57.5%; Score 23; DB 3; Length 2634;
Best Local Similarity 74.4%; Pred. No. 0.81;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 2 tcgagccaccataaccctcaactccaggattggg 40
||| ||||| || ||||| ||||| ||||| |||||
Db 2056 TCCACCCACACTTGAAGCCACAAAACTCAGGATTGGG 2018

RESULT 4
US-09-115-934A-2
; Sequence 2, Application US/09115934A
; Patent No. 6130317
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Xu, Qunli
; TITLE OF INVENTION: BAX Inhibitor Proteins
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/115,934A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/818,514
; FILING DATE: 14-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3209
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2634 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-115-934A-2
```


Db 2326 AGCTCCCTCACCCTGGATCCACGAGGACAGGG 2291

RESULT 10

US-09-221-235-10
; Sequence 10, Application US/09221235
; Patent No. 6043040
; GENERAL INFORMATION:

; APPLICANT: Acton, Susan
; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
; FILE REFERENCE: MNI-050
; CURRENT APPLICATION NUMBER: US/09/221,235
; CURRENT FILING DATE: 1998-12-28
; EARLIER APPLICATION NUMBER: 09/163,115
; EARLIER FILING DATE:
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1864
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (275)..(754)
US-09-221-235-10

Query Match 48.0%; Score 19.2; DB 3; Length 1864;
Best Local Similarity 67.5%; Pred. No. 29;
Matches 27; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 ctcgagccaccataaccctcaatactccaggattggg 40
||| ||| |||| | ||| |||| | ||||
Db 955 ctcagcattcccaagctcttaattctccataaaatggg 994

RESULT 11

US-09-221-928-10
; Sequence 10, Application US/09221928
; Patent No. 6121030
; GENERAL INFORMATION:

; APPLICANT: Acton, Susan
; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
; FILE REFERENCE: MNI-050
; CURRENT APPLICATION NUMBER: US/09/221,928
; CURRENT FILING DATE: 1998-12-28
; EARLIER APPLICATION NUMBER: 09/163,115
; EARLIER FILING DATE:
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1864
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (275)..(754)
US-09-221-928-10

Query Match 48.0%; Score 19.2; DB 3; Length 1864;
Best Local Similarity 67.5%; Pred. No. 29;
Matches 27; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 ctcgagccaccataaccctcaatactccaggattggg 40
||| ||| |||| | ||| |||| | ||||
Db 955 ctcagcattcccaagctcttaattctccataaaatggg 994

RESULT 12

US-09-221-527-10
; Sequence 10, Application US/09221527
; Patent No. 6146832

; GENERAL INFORMATION:
; APPLICANT: Acton, Susan
; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
; FILE REFERENCE: MNI-050
; CURRENT APPLICATION NUMBER: US/09/221,527
; CURRENT FILING DATE: 1998-12-28
; EARLIER APPLICATION NUMBER: 09/163,115
; EARLIER FILING DATE:
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1864
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (275)..(754)
US-09-221-527-10

Query Match 48.0%; Score 19.2; DB 3; Length 1864;
Best Local Similarity 67.5%; Pred. No. 29;
Matches 27; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 ctcgagccaccataaccctcaatactccaggattggg 40
||| ||| |||| | ||| |||| | ||||
Db 955 ctcagcattcccaagctcttaattctccataaaatggg 994

RESULT 13

US-09-221-236-10
; Sequence 10, Application US/09221236
; Patent No. 6146841
; GENERAL INFORMATION:

; APPLICANT: Acton, Susan
; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
; FILE REFERENCE: MNI-050
; CURRENT APPLICATION NUMBER: US/09/221,236
; CURRENT FILING DATE: 1998-12-28
; EARLIER APPLICATION NUMBER: 09/163,115
; EARLIER FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1864
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (275)..(754)
US-09-221-236-10

Query Match 48.0%; Score 19.2; DB 3; Length 1864;
Best Local Similarity 67.5%; Pred. No. 29;
Matches 27; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 ctcgagccaccataaccctcaatactccaggattggg 40
||| ||| |||| | ||| |||| | ||||
Db 955 ctcagcattcccaagctcttaattctccataaaatggg 994

RESULT 14

US-09-221-416-10
; Sequence 10, Application US/09221416
; Patent No. 6153417
; GENERAL INFORMATION:

; APPLICANT: Acton, Susan
; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
; FILE REFERENCE: MNI-050
; CURRENT APPLICATION NUMBER: US/09/221,416
; CURRENT FILING DATE: 1998-12-28
; EARLIER APPLICATION NUMBER: 09/163,115

; EARLIER FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1864
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (275)..(754)
US-09-221-416-10

Query Match 48.0%; Score 19.2; DB 3; Length 1864;
Best Local Similarity 67.5%; Pred. No. 29;
Matches 27; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 1 ctccagccaccataaccctcaatactaccagggtggg 40
||| ||| |||| | ||| ||| |||| | ||||
Db 955 ctccagcattcccaaaagctcttaattctccataaaatggg 994

RESULT 15
US-09-221-245-10
; Sequence 10, Application US/09221245
; Patent No. 6180358
; GENERAL INFORMATION:
; APPLICANT: Acton, Susan
; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
; FILE REFERENCE: MNI-050
; CURRENT APPLICATION NUMBER: US/09/221,245
; CURRENT FILING DATE: 1998-12-28
; EARLIER APPLICATION NUMBER: US 09/163,115
; EARLIER FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1864
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (275)..(754)
US-09-221-245-10

Query Match 48.0%; Score 19.2; DB 4; Length 1864;
Best Local Similarity 67.5%; Pred. No. 29;
Matches 27; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 1 ctccagccaccataaccctcaatactaccagggtggg 40
||| ||| |||| | ||| ||| |||| | ||||
Db 955 ctccagcattcccaaaagctcttaattctccataaaatggg 994

Search completed: June 4, 2002, 17:47:11
Job time: 6584 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 17:56:12 ; Search time 521.5 Seconds
(without alignments)
131.690 Million cell updates/sec

Title: US-09-721-543A-13
Perfect score: 40
Sequence: 1 ctcgagccaccataacccatcaatactccaggattggg 40

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*			
1:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*		
2:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*		
3:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*		
4:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*		
5:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*		
6:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*		
7:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*		
8:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*		
9:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*		
10:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*		
11:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*		
12:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*		
13:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*		
14:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*		
15:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*		
16:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*		
17:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*		
18:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*		
19:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*		
20:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*		
21:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*		
22:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*		
23:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*		
24:	/SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match	Description
1	40	100.0	40 22 AAD10599
2	23	57.5	2634 19 AAV59067
3	23	57.5	2922 21 AAF18054
4	21.2	53.0	121 22 ABA71669
5	21.2	53.0	121 22 ABA37784
6	21.2	53.0	121 22 AAK20013
7	21.2	53.0	121 22 AAK46055
8	21.2	53.0	121 22 AAI51965
9	21.2	53.0	473 22 ABA59145

c 10	21.2	53.0	473	22	ABA27922	Probe #6388 for ge
c 11	21.2	53.0	473	22	AAK07330	Human brain expres
c 12	21.2	53.0	473	22	AAK33095	Human bone marrow
c 13	21.2	53.0	473	22	AAI38893	Probe #7579 used t
c 14	20.8	52.0	837	20	AAK37525	Human secreted pro
c 15	20.8	52.0	3896	23	ABL06092	Drosophila melanog
c 16	20.6	51.5	366	22	AAI87560	Human polynucleoti
c 17	20.6	51.5	744	22	AAI96641	Human neuroblastom
c 18	20.6	51.5	1907	22	AAD08058	Human extracellular
c 19	20.4	51.0	5592	22	AAF30935	Spodoptera frugipe
c 20	20	50.0	897	23	AAK81288	DNA encoding novel
c 21	20	50.0	1378	23	AAK81354	DNA encoding novel
c 22	20	50.0	3364	19	AAV07266	Muscle creatine ki
c 23	20	50.0	16217	24	ABL32625	Human immune syste
c 24	19.8	49.5	2832	15	AAQ62186	Cyanobacteria repl
c 25	19.8	49.5	4809	15	AAQ62185	pAQ1 plasmid fragm
c 26	19.8	49.5	15196	22	AAK73103	Human immune/haema
c 27	19.8	49.5	15196	22	AAK87548	Human immune/haema
c 28	19.8	49.5	32249	22	ABA20005	Human nervous syst
c 29	19.8	49.5	32249	22	ABA07406	Human pancreatic c
c 30	19.8	49.5	32249	22	AAK91137	Human digestive sy
c 31	19.8	49.5	50442	22	AAK73083	Human immune/haema
c 32	19.8	49.5	50442	22	AAK87551	Human immune/haema
c 33	19.6	49.0	511	23	AAK80311	DNA encoding novel
c 34	19.6	49.0	791	22	AAI91505	Human polynucleoti
c 35	19.6	49.0	32591	22	AAK43104	Human Oestrogen re
c 36	19.4	48.5	282	21	AAA43173	Xenopus secreted e
c 37	19.4	48.5	476	22	ABA57621	Human foetal liver
c 38	19.4	48.5	476	22	AAK05574	Human brain expres
c 39	19.4	48.5	476	22	AAK31288	Human bone marrow
c 40	19.4	48.5	476	22	AAI37184	Probe #5870 used t
c 41	19.4	48.5	674	20	AAK37371	Human secreted pro
c 42	19.4	48.5	1554	20	AAK16575	Xenopus WA545 prot
c 43	19.4	48.5	1617	23	AAK85386	DNA encoding novel
c 44	19.4	48.5	12789	24	AAI16376	Human interleukin
c 45	19.4	48.5	34980	22	AAH64366	C glutamicum codin

ALIGNMENTS

RESULT 1	
AD	AAAD10599 standard; DNA; 40 BP.
XX	
AC	AAAD10599;
XX	
DT	24-SEP-2001 (first entry)
XX	
DE	DNA ligand #5 for human cytomegalovirus.
XX	
KW	Ligand; human cytomegalovirus; HCMV infection; herpetic ulcer;
KW	pneumonia; gastroenteritis; chorioretinitis; antiviral agent;
KW	herpesvirus infection; lesion; ds.
XX	
OS	Human cytomegalovirus.
OS	Synthetic.
PN	WO200138341-A1.
XX	
PD	31-MAY-2001.
XX	
PF	21-NOV-2000; 2000WO-US32259.
XX	
PR	24-NOV-1999; 99US-0167509.
XX	
PA	(REGC) UNIV CALIFORNIA.
XX	
PI	Liu F, Wang J, Jiang H;
XX	
DR	WPI; 2001-367658/38.
XX	
PT	Polynucleotide ligands useful as anti-viral agents for the treatment of

PT herpesvirus infections e.g. primary or chronic infections with human
PT cytomegalovirus

XX Claim 8; Page 6; 60pp; English.

XX The invention relates to polynucleotide ligands which bind to
CC herpesvirus, particularly human cytomegalovirus (HCMV) to decrease
CC or block HCMV infection in target cells. These ligands are used as
CC anti-viral agents for the treatment of herpesvirus infections (e.g.
CC lesions and herpetic ulcers) and particularly HCMV associated
CC primary or chronic infections such as pneumonia, gastroenteritis
CC and chorioretinitis. The present sequence is DNA ligand which bind
CC to HCMV to decrease or block HCMV infection in target cells.

XX Sequence 40 BP; 10 A; 15 C; 8 G; 7 T; 0 other;

Query Match 100.0%; Score 40; DB 22; Length 40;

Best Local Similarity 100.0%; Pred. No. 1.1e-07; Length 40;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctcgagccaccataaacctcaatactccaggattggg 40

Db 1 ctcgagccaccataaacctcaatactccaggattggg 40

RESULT 2

AAV59067/c
ID AAV59067 standard; cDNA; 2634 BP.

AC AAV59067;

DT 02-FEB-1999 (first entry)

XX Bax inhibitor BI-1 cDNA.

XX Bax inhibitor; BI-1; human; apoptosis; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 73..786

FT /*tag= a

PN WO9840397-A1.

XX 17-SEP-1998.

XX 13-MAR-1998; 98WO-US05015.

XX 14-MAR-1997; 97US-0818514.

XX (BURN-) BURNHAM INST.

XX Reed JC, Xu Q;

XX WPI; 1998-531519/45.

XX P-PSDB; AAW73136.

PT Bax inhibitor proteins, BI-1 and BI-2 - useful e.g. to modulate
PT cellular apoptotic activity or identify agents altering BI-1 or BI-2
PT binding which can modulate apoptotic activity

XX Claim 2; Page 61-63; 80pp; English.

XX This cDNA clone codes for an inhibitor protein, termed BI-1 (see
CC AAW73136), of the pro-apoptotic protein Bax. Nucleic acids encoding
CC BI-1 and BI-2 (see AAV59068) were identified by suppression of
CC Bax-induced death of yeast cells transformed to express human Bax.
CC A human HepG2 cDNA library was used for library screening. The
CC invention provides vectors, optionally expression or viral vectors,
CC containing BI nucleic acids, and host cells containing these
CC vectors. The nucleic acids encoding BI-1/BI-2 can be used to

CC increase expression of these proteins in cells, or antisense
CC molecules prepared from them used to decrease expression. In
CC these ways, cellular apoptotic activity may be modulated (claimed).
CC The nucleic acids and complementary sequences are also useful as
CC probes to detect BI-encoding nucleic acid molecules in samples.

XX Sequence 2634 BP; 632 A; 626 C; 597 G; 779 T; 0 other;

Query Match 57.5%; Score 23; DB 19; Length 2634;

Best Local Similarity 74.4%; Pred. No. 4.1;

Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 tcgagccaccataaacctcaatactccaggattggg 40

Db 2056 TCCACCCACACTGTAAGCCACAACTGCAGGGATTGGG 2018

RESULT 3

AAF18054/c

ID AAF18054 standard; DNA; 2922 BP.

AC AAF18054;

DT 14-MAR-2001 (first entry)

XX Lung cancer associated polynucleotide sequence SEQ ID 73.

XX Human; lung cancer associated protein; neuroprotective; cytostatic;
KW cardioactive; immunomodulatory; muscular active; vulnery;
KW gastrointestinal; nephrotropic; antiinfective; gynecological;
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW proliferative disorder; wound healing; infectious disease; ds.

OS Homo sapiens.

XX WO200055180-A2.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05918.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (ROSE/) ROSEN C A.

XX Ruben SM;

XX WPI; 2000-587514/55.

XX P-PSDB; AAB58178.

XX Lung cancer associated gene sequences, referred to as lung cancer
XX antigens, useful for treatment, prevention, and diagnosis of disorders
XX such as lung cancer -

XX Claim 1; Page 550-551; 1425pp; English.

XX Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective; cytostatic; cardioactive;
CC immunomodulatory; muscular active general; vulnery; gastrointestinal
CC general; nephrotropic; antiinfective; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the
CC protein or polynucleotide sequences. The lung cancer associated
CC polynucleotide sequences may be used for detection of lung cancer,
CC chromosome identification, as chromosome markers, and for numerous other
CC diagnostic or research purposes. The proteins may be used to treat
CC disorders such as neural, immune, muscular, reproductive
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders. The proteins may also be used in the treatment of wounds and
CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and

Qy 7 ccaccataaccctcaatactccaggaattggg 40

AAK20013/C
ID AAK20013 standard: DNA: 121 BP

PA (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-488897/53.
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
PT
PS Claim 25; SEQ ID No 20651; 654pp; English.
XX The present invention relates to single exon nucleic acid probes (SENP).
XX The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
XX Sequence 121 BP; 46 A; 20 C; 28 G; 27 T; 0 other;
SQ

Query Match 53.0%; Score 21.2; DB 22; Length 121;
Best Local Similarity 76.5%; Pred. No. 14;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
Qy 7 ccacccataaacctcaataactccaggattggg 40
Db 38 CCTTCTATAATCCTCCTACTTACTCAAGGTAATGGG 5

RESULT 9
ABA59145/C
ID ABA59145 standard; DNA; 473 BP.
XX AC ABA59145;
XX
XX 01-FEB-2002 (first entry)
XX
XX Human foetal liver single exon nucleic acid probe #7450.
XX
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
XX Homo sapiens.
XX
XX WO200157277-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00669.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver -
PT
PS Claim 1; SEQ ID NO 7450; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid

CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 473 BP; 144 A; 91 C; 95 G; 143 T; 0 other;
SQ

Query Match 53.0%; Score 21.2; DB 22; Length 473;
Best Local Similarity 76.5%; Pred. No. 18;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
Qy 7 ccacccataaacctcaataactccaggattggg 40
Db 347 CCTTCTATAATCCTCCTACTTACTCAAGGTAATGGG 314

RESULT 10
ABA27922/C
ID ABA27922 standard; DNA; 473 BP.
XX AC ABA27922;
XX
XX 23-JAN-2002 (first entry)
XX
XX Probe #6388 for gene expression analysis in human heart cell sample.
XX
XX Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
XX Homo sapiens.
XX
XX WO200157274-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00666.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488899/53.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts -
PT
PS Claim 1; SEQ ID No 6388; 530pp; English.
XX
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart. The
XX present sequence is one such probe. The probes may be used for
XX predicting, measuring and displaying gene expression in samples derived
XX from the human heart via microarrays. By measuring gene expression, the
XX probes are useful for predicting, diagnosing, grading, staging,
XX monitoring and prognosing diseases of the human heart and vascular system
XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
XX congenital heart disease.
XX
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 473 BP; 144 A; 91 C; 95 G; 143 T; 0 other;
SQ

Query Match 53.0%; Score 21.2; DB 22; Length 473;
Best Local Similarity 76.5%; Pred. No. 18;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 7 ccacccataaccctcaataactccaggattggg 40
Db 347 CCTTCTATAATCCTCAATTAAGGTAATGGG 314

RESULT 11

AAK07330/c
ID AAK07330 standard; DNA; 473 BP.

XX AAK07330;

XX AAK07330;

XX 05-NOV-2001 (first entry)

XX Human brain expressed single exon probe SEQ ID NO: 7321.

XX Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.

XX Homo sapiens.

OS WO200157275-A2.

PN 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00667.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX Example 4; SEQ ID NO: 7321; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.

XX Sequence 473 BP; 144 A; 91 C; 95 G; 143 T; 0 other;

Query Match 53.0%; Score 21.2; DB 22; Length 473;
Best Local Similarity 76.5%; Pred. No. 18;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 7 ccacccataaccctcaataactccaggattggg 40
Db 347 CCTTCTATAATCCTCAATTAAGGTAATGGG 314

RESULT 12

AAK33095/c

ID AAK33095 standard; DNA; 473 BP.
XX AAK33095;
XX 06-NOV-2001 (first entry)
XX Human bone marrow expressed single exon probe SEQ ID NO: 7652.
DE Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX Homo sapiens.
XX WO200157276-A2.
XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00668.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488900/53.

XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX Example 4; SEQ ID NO: 7652; 658pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.

XX Sequence 473 BP; 144 A; 91 C; 95 G; 143 T; 0 other;

Query Match 53.0%; Score 21.2; DB 22; Length 473;
Best Local Similarity 76.5%; Pred. No. 18;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 7 ccacccataaccctcaataactccaggattggg 40
Db 347 CCTTCTATAATCCTCAATTAAGGTAATGGG 314

RESULT 13

AAI38893/c

ID AAI38893 standard; DNA; 473 BP.

XX AAI38893;

XX 17-OCT-2001 (first entry)

XX Probe #7579 used to measure gene expression in human placenta sample.
DE Probe; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder; ss.
XX Homo sapiens.

OS WO200157272-A2.

PN


```
XX 09-AUG-2001.
PD
XX 30-JAN-2001; 2001WO-US00663.
PF
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
PT
XX Claim 25; SEQ ID No 7579; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
XX Sequence 473 BP; 144 A; 91 C; 95 G; 143 T; 0 other;
SQ

Query Match 53.0%; Score 21.2; DB 22; Length 473;
Best Local Similarity 76.5%; Pred. No. 18;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 7 ccaccataaccctcaatactccaggattggg 40
DB 347 CCTTCCTATTAATCTCATTACTCAAGGTATGGG 314

RESULT 14
AA37525/c
ID NAX37525 standard; cDNA; 837 BP.
XX
XX NAX37525;
XX
XX 06-JUL-1999 (first entry)
XX
XX Human secreted protein cDNA fragment containing gene 75.
XX
XX Human; secreted protein; treatment; prevention; protein therapy; AIDS;
XX gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
XX developmental abnormality; fetal deficiency; blood disease; leukemia;
XX immune system disease; autoimmune disease; hepatic disease; lymphoma;
XX renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
XX cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
XX pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
XX arthritis; malignancy; digestive; endocrine; infection; ss.
XX
XX Homo sapiens.
XX
XX WO9918208-A1.
XX
XX 15-APR-1999.
XX
XX 01-OCT-1998; 98WO-US20775.
XX
XX 02-OCT-1997; 97US-0060884.
XX
XX 02-OCT-1997; 97US-0060833.
XX
XX 02-OCT-1997; 97US-0060836.
XX
XX 02-OCT-1997; 97US-0060837.
XX

PR 02-OCT-1997; 97US-0060838.
PR 02-OCT-1997; 97US-0060839.
PR 02-OCT-1997; 97US-0060843.
PR 02-OCT-1997; 97US-0060862.
PR 02-OCT-1997; 97US-0060866.
PR 02-OCT-1997; 97US-0060874.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Carter KC, Duan DR, Endress GA, Feng P, Ferrie AM;
XX Florence KA, Greene JM, Janat F, Lafleur DW, NI J;
XX Rosen CA, Ruben SM, Shi Y, Young P, Yu G;
XX
XX WPI; 1999-264022/22.
XX
XX P-PSDB; AAY07926.
XX
XX New isolated human genes and the secreted polypeptides they encode
XX
XX Claim 1a; Page 264; 368pp; English.
XX
XX This invention describes novel isolated human genes and the secreted
XX proteins they encode. The products of the invention are useful for
XX preventing, treating or ameliorating medical conditions, e.g. by protein
XX or gene therapy. Also pathological conditions can be diagnosed by
XX determining the amount of the new polypeptides in a sample or by
XX determining the presence of mutations in the new polynucleotides.
XX Specific uses are described for each of the 101 polynucleotides,
XX on which tissues they are most highly expressed in, and include
XX developing products for the diagnosis or treatment of cancer, tumours,
XX neurodegenerative disorders, developmental abnormalities and fetal
XX deficiencies, blood disorders, leukemias, diseases of the immune system,
XX autoimmune diseases, hepatic and renal disease, lymphomas, inflammation,
XX allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate
XX disease, skeletal or cardiac muscle disorders, pulmonary disorders,
XX transplant rejection, disorders involving osteoclasts such as
XX osteoporosis, arthritis or malignancies, digestive/endocrine disorders,
XX infections and AIDS. The human secreted proteins of the invention are
XX represented in AAY07852-Y07993 and the encoding nucleic acids are
XX represented in AAX37451-X37552.
XX
XX Sequence 837 BP; 171 A; 249 C; 187 G; 228 T; 2 other;
SQ

Query Match 52.0%; Score 20.8; DB 20; Length 837;
Best Local Similarity 70.0%; Pred. No. 29;
Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 ctgagccaccataaccctcaatactccaggattggg 40
DB 250 CTCAGGCGCCCGAGACACTCCAGACGCTGGAGATTGGG 211

RESULT 15
ABL06092
ID ABL06092 standard; cDNA; 3896 BP.
XX
XX ABL06092;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 12758.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
```

PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI: 2001-656860/75.
 DR P-PSDB; ABB61989.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1: SEQ ID NO 12758; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 3896 BP; 1147 A; 931 C; 883 G; 935 T; 0 other;

Query Match 52.0%; Score 20.8; DB 23; Length 3896;
 Best Local Similarity 70.0%; Pred. No. 38;
 Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 1 ctcgagccaccataaccctcaatactccagggtggg 40
 ||||| ||||| || ||||| |||||
 Db 3333 ctcgaagcaccacacacagcagcacgctccacggtggg 3372

Search completed: June 4, 2002, 17:56:14
 Job time: 7102 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 17:45:41 ; Search time 2137.99 Seconds
(without alignments)
391.518 Million cell updates/sec

Title: US-09-721-543A-13
Perfect score: 40
Sequence: 1 ctgagccacccataaccctcaatactccaggattggg 40

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_hgt.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pi.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vl.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pi.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vl.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htgo_inv.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
------------	-------	-------	-------	--------	----	----	-------------

RESULT	1
LOCUS	AK026605/c
DEFINITION	Homo sapiens cDNA: FLJ22952 fis, clone KAT09742.
ACCESSION	AK026605
VERSION	AK026605.1 GI:10439494
KEYWORDS	oligo capping; fis (full insert sequence).
SOURCE	Homo sapiens signet-ring cell carcinoma cell_line:KATO III CDNA to mRNA, clone_lib:KAT clone:KAT09742.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (sites) Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
TITLE	NEDO human cDNA sequencing project
JOURNAL	Unpublished (2000)
REFERENCE	2 (bases 1 to 1653) Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.
AUTHORS	Direct Submission
TITLE	Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
JOURNAL	

ALIGNMENTS

C	1	23	57.5	1653	9	AK026605	AK026605 Homo sapi
C	2	23	57.5	2600	9	HSTEGT	X75861 H.sapiens T
C	3	23	57.5	2609	9	BC000916	BC000916 Homo sapi
C	4	23	57.5	2634	6	AR058920	AR058920 Sequence
C	5	23	57.5	2634	6	AR058921	AR058921 Sequence
C	6	23	57.5	2634	6	AR112791	AR112791 Sequence
C	7	23	57.5	2634	6	AR112792	AR112792 Sequence
C	8	23	57.5	197189	2	AC019168	AC019168 Homo sapi
C	9	23	57.5	215867	2	AC084037	AC084037 Homo sapi
C	10	23	57.5	225818	2	AC090998	AC090998 Homo sapi
C	11	22.8	57.0	171185	2	AP002424	AP002424 Homo sapi
C	12	22.8	57.0	177097	2	AP001569	AP001569 Homo sapi
C	13	22.8	57.0	179726	9	AC007052	AC007052 Homo sapi
C	14	22.8	57.0	182411	2	AC090408	AC090408 Homo sapi
C	15	22.8	57.0	188357	9	AL159996	AL159996 Human DNA
C	16	22.8	57.0	189329	2	AC090230	AC090230 Homo sapi
C	17	22.8	57.0	196869	2	AC087535	AC087535 Homo sapi
C	18	22.8	57.0	200774	2	AP001592	AP001592 Homo sapi
C	19	22.4	56.0	137845	2	AC010367	AC010367 Homo sapi
C	20	22.4	56.0	151514	9	HS471M13	Z97198 Human DNA s
C	21	22	55.0	452	11	G22123	G22123 human STS W
C	22	22	55.0	34887	3	TBR012199	AJ012199 Trypanoso
C	23	22	55.0	76727	9	HS821D11	AL021453 Human DNA
C	24	22	55.0	112022	9	HSJ9E11	AJ009611 Homo sapi
C	25	22	55.0	177540	9	AC006538	AC006538 Homo sapi
C	26	21.8	54.5	145607	2	AC105341	AC105341 Homo sapi
C	27	21.8	54.5	152339	2	AC083931	AC083931 Homo sapi
C	28	21.8	54.5	154165	2	AC087620	AC087620 Homo sapi
C	29	21.8	54.5	154599	2	AC016997	AC016997 Homo sapi
C	30	21.8	54.5	157021	2	AC020991	AC020991 Homo sapi
C	31	21.6	54.0	36415	9	AC093645	AC093645 Homo sapi
C	32	21.6	54.0	37470	9	CNS00YVD	AL036810 Homo sapi
C	33	21.6	54.0	110000	9	HS611N7	AL035663 Human DNA
C	34	21.6	54.0	104436	2	LMFLCHR12_4	Continuation (5 of
C	35	21.6	54.0	122952	2	AL513483	AL513483 Homo sapi
C	36	21.6	54.0	173855	2	OSN00181	AL662983 Oryza sat
C	37	21.6	54.0	210442	2	AL445199	AL445199 Homo sapi
C	38	21.6	54.0	226860	9	AL359920	AL359920 Human DNA
C	39	21.4	53.5	54525	2	AC103846	AC103846 Homo sapi
C	40	21.4	53.5	70993	2	AC102199	AC102199 Mus muscu
C	41	21.4	53.5	161820	2	AC095091	AC095091 Rattus no
C	42	21.4	53.5	217345	2	AC104393	AC104393 Homo sapi
C	43	21.2	53.0	65824	2	AC099899	AC099899 Mus muscu
C	44	21.2	53.0	70594	2	AC015786	AC015786 Homo sapi
C	45	21.2	53.0	116252	9	AL445228	AL445228 Human DNA


```

BASE COUNT      606 a      625 c      602 g      776 t
ORIGIN
RKLMLAMNEKDKKKKK"

Query Match      57.5%; Score 23; DB 9; Length 2609;
Best Local Similarity 74.4%; Pred. No. 31;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 tcgagcccccataacctcaatactccaggattggg 40
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 2049 TCCACCCACACTTGAAGCCACAAACTGCAGGATTGGG 2011

RESULT 4
LOCUS      AR058920/c      2634 bp      DNA      linear      PAT 29-SEP-1999
DEFINITION Sequence 1 from patent US 5837838.
ACCESSION  AR058920
VERSION     AR058920.1 GI:5984497
KEYWORDS
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE   1 (bases 1 to 2634)
AUTHORS    Reed,J.C. and Xu,Q.
TITLE      Bax inhibitor proteins
JOURNAL    Patent: US 5837838-A 1 17-NOV-1998;
FEATURES   Location/Qualifiers
            source
            1..2634
            /organism="unknown"

BASE COUNT      632 a      626 c      597 g      779 t
ORIGIN

Query Match      57.5%; Score 23; DB 6; Length 2634;
Best Local Similarity 74.4%; Pred. No. 31;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 tcgagcccccataacctcaatactccaggattggg 40
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 2056 TCCACCCACACTTGAAGCCACAAACTGCAGGATTGGG 2018

RESULT 5
LOCUS      AR058921      2634 bp      DNA      linear      PAT 29-SEP-1999
DEFINITION Sequence 2 from patent US 5837838.
ACCESSION  AR058921
VERSION     AR058921.1 GI:5984498
KEYWORDS
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE   1 (bases 1 to 2634)
AUTHORS    Reed,J.C. and Xu,Q.
TITLE      Bax inhibitor proteins
JOURNAL    Patent: US 5837838-A 2 17-NOV-1998;
FEATURES   Location/Qualifiers
            source
            1..2634
            /organism="unknown"

BASE COUNT      779 a      597 c      626 g      632 t
ORIGIN

Query Match      57.5%; Score 23; DB 6; Length 2634;
Best Local Similarity 74.4%; Pred. No. 31;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 tcgagcccccataacctcaatactccaggattggg 40
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 2056 TCCACCCACACTTGAAGCCACAAACTGCAGGATTGGG 2018

RESULT 6
LOCUS      AR112791/c      2634 bp      DNA      linear      PAT 16-MAY-2001
DEFINITION Sequence 1 from patent US 6130317.
ACCESSION  AR112791
VERSION     AR112791.1 GI:14092691
KEYWORDS
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE   1 (bases 1 to 2634)
AUTHORS    Reed,J.C. and Xu,Q.
TITLE      Bax inhibitor proteins
JOURNAL    Patent: US 6130317-A 1 10-OCT-2000;
FEATURES   Location/Qualifiers
            source
            1..2634
            /organism="unknown"

BASE COUNT      632 a      626 c      597 g      779 t
ORIGIN

Query Match      57.5%; Score 23; DB 6; Length 2634;
Best Local Similarity 74.4%; Pred. No. 31;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 tcgagcccccataacctcaatactccaggattggg 40
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 2056 TCCACCCACACTTGAAGCCACAAACTGCAGGATTGGG 2018

RESULT 7
LOCUS      AR112792      2634 bp      DNA      linear      PAT 16-MAY-2001
DEFINITION Sequence 2 from patent US 6130317.
ACCESSION  AR112792
VERSION     AR112792.1 GI:14092692
KEYWORDS
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE   1 (bases 1 to 2634)
AUTHORS    Reed,J.C. and Xu,Q.
TITLE      Bax inhibitor proteins
JOURNAL    Patent: US 6130317-A 2 10-OCT-2000;
FEATURES   Location/Qualifiers
            source
            1..2634
            /organism="unknown"

BASE COUNT      779 a      597 c      626 g      632 t
ORIGIN

Query Match      57.5%; Score 23; DB 6; Length 2634;
Best Local Similarity 74.4%; Pred. No. 31;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 tcgagcccccataacctcaatactccaggattggg 40
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 579 TCCACCCACACTTGAAGCCACAAACTGCAGGATTGGG 617

RESULT 8
LOCUS      AC019168      197189 bp      DNA      linear      HTG 07-JUL-2000
DEFINITION Homo sapiens chromosome 12 clone RP11-161E16, WORKING DRAFT
ACCESSION  AC019168
VERSION     AC019168.3 GI:7230912
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 197189)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 197189)
Waterston,R.H.
Direct Submission
Submitted (30-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Mar 13, 2000 this sequence version replaced gi:7021812.

COMMENT

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
----- Project Information -----
Center project name: H.NH0161E16
----- Summary Statistics -----
Sequencing vector: M13; 92%
Sequencing vector: plasmid; 8%
Chemistry: Dye-terminator; 92% of reads
Chemistry: Dye-terminator Big Dye; 8% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 180505 bases at least Q40
Consensus quality: 186042 bases at least Q30
Consensus quality: 189445 bases at least Q20
Insert size: 190000; agarose-fp
Insert size: 194889; sum-of-contigs
Quality coverage: 3.69 in Q20 bases; agarose-fp
Quality coverage: 3.65 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1451: contig of 1451 bp in length
1551: gap of unknown length
1552: contig of 1355 bp in length
2907: gap of unknown length
3006: contig of 1298 bp in length
4305: gap of unknown length
4405: contig of 1217 bp in length
5622: gap of unknown length
5721: contig of 2378 bp in length
8099: gap of unknown length
8100: contig of 2601 bp in length
8200: gap of unknown length
10801: contig of 3300 bp in length
10901: gap of unknown length
14201: contig of 3271 bp in length
14301: gap of unknown length
17572: contig of 4185 bp in length
17672: gap of unknown length
21857: gap of unknown length
21957: contig of 3792 bp in length
25749: gap of unknown length
25849: contig of 3994 bp in length
29843: gap of unknown length
29943: contig of 5479 bp in length
35422: gap of unknown length
35522: contig of 5453 bp in length
40975: gap of unknown length
41075: contig of 7517 bp in length
48591: gap of unknown length
48592: contig of 9477 bp in length
48692: gap of unknown length
58169: contig of 9764 bp in length
58269: gap of unknown length
68032: contig of 9331 bp in length
68133: gap of unknown length
77463: contig of 9331 bp in length

* 77464 77563: gap of unknown length
* 77564 93029: contig of 15466 bp in length
* 93030 93129: gap of unknown length
* 93130 107954: contig of 14825 bp in length
* 107955 108054: gap of unknown length
* 108055 121842: contig of 13788 bp in length
* 121843 121942: gap of unknown length
* 121943 137362: contig of 15420 bp in length
* 137363 137462: gap of unknown length
* 137463 155524: contig of 18062 bp in length
* 155525 155624: gap of unknown length
* 155625 173416: contig of 17792 bp in length
* 173417 173516: gap of unknown length
* 173517 197189: contig of 23673 bp in length.
FEATURES
Location/Qualifiers
1..197189
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-161E16"
1..1451
/note="assembly_name:Contig3"
1552..2906
/note="assembly_name:Contig4"
3007..4304
/note="assembly_name:Contig5"
4405..5621
/note="assembly_name:Contig6"
5722..8099
/note="assembly_name:Contig7"
8200..10800
/note="assembly_name:Contig8"
10901..14200
/note="assembly_name:Contig9"
14301..17571
/note="assembly_name:Contig10"
17672..21856
/note="assembly_name:Contig11"
21957..25748
/note="assembly_name:Contig12"
25849..29842
/note="assembly_name:Contig13"
29943..35421
/note="assembly_name:Contig14"
35522..40974
/note="assembly_name:Contig15"
41075..48591
/note="assembly_name:Contig16"
48692..58168
/note="assembly_name:Contig17"
58269..68032
/note="assembly_name:Contig18"
68133..77463
/note="assembly_name:Contig19"
77564..93029
/note="assembly_name:Contig20"
93130..107954
/note="assembly_name:Contig21"
clone_end:SP6
vector_side:right
108055..121842
/note="assembly_name:Contig22"
121943..137362
/note="assembly_name:Contig23"
137463..155524
/note="assembly_name:Contig24"
155625..173416
/note="assembly_name:Contig25"
173517..197189
/note="assembly_name:Contig26"
BASE COUNT 51312 a 47093 c 47437 g 49028 t 2319 others
ORIGIN

Query Match	57.5%	Score 23	DB 2	Length 197189
Best Local Similarity	74.4%	Pred. No. 27		
Matches	29	Conservative	0	Mismatches 10; Indels 0; Gaps 0

QY	2	tcagccacccataccctcaatactaccaggattggg 40
Db	123952	TCCACCCACACTTGAACCCACAAACTGCAGGGATTGGG 123990

RESULT	9
AC084037	
LOCUS	AC084037
DEFINITION	Homo sapiens chromosome 12q clone RP11-382G3, WORKING DRAFT SEQUENCE, 7 unordered pieces.
ACCESSION	AC084037
VERSION	AC084037.28
KEYWORDS	HTG: HTGS_PASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE	human.
ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE	1 (bases 1 to 215867)
AUTHORS	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbara,J., Benton,J., Bimge,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burrell,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Correll,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lied,C., Liu,J., Liu,W., Loulseged,H., Lozadow,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,A., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rivers,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,Y., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G., and Gibbs,R.
TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 215867)
AUTHORS	Worley,K.C.
TITLE	Direct Submission
JOURNAL	Submitted (11-OCT-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT	On Jan 31, 2002 this sequence version replaced gi:15809132. ----- Genome Center

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AP002424
AP002424.1 GI:8131588
HTG: HTGS_PHASE1; HTGS_DRAFT.
Homo sapiens DNA, clone:RP11-873L22.
Homo sapiens

REFERENCE

1 (bases 1 to 171185)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE

Homo sapiens 171,185 genomic DNA of 18q21

REFERENCE

2 (bases 1 to 171185)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

AUTHORS

Submitted (29-MAY-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan [E-mail:hattori@gsc.riken.go.jp].
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)

JOURNAL

Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: HumDraft18
Center clone name: RP11-873L22
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 156977 bases at least Q40
Consensus quality: 164110 bases at least Q30
Consensus quality: 167421 bases at least Q20
Insert size: 169485; sum-of-contigs
Quality coverage: 4.80x in Q20 bases; sum-of-contigs

COMMENT

NOTE: This is a 'working draft' sequence. It currently consists of
18 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved

1 24923 contig of 24923 bp in length
25024 47004 contig of 21981 bp in length
47105 67070 contig of 19966 bp in length
82858 82757 contig of 15887 bp in length
95645 95544 contig of 12687 bp in length
109693 109592 contig of 13948 bp in length
120268 120167 contig of 10475 bp in length
129356 129255 contig of 8988 bp in length
137104 144376 contig of 7648 bp in length
150760 150659 contig of 6183 bp in length
160623 163263 contig of 5137 bp in length
165295 167704 contig of 4526 bp in length
167805 167704 contig of 2641 bp in length
169851 171185 contig of 1831 bp in length
169851 171185 contig of 1335 bp in length

1 24923 contig of 24923 bp in length
25024 47004 contig of 21981 bp in length
47105 67070 contig of 19966 bp in length
82858 82757 contig of 15887 bp in length
95645 95544 contig of 12687 bp in length
109693 109592 contig of 13948 bp in length
120268 120167 contig of 10475 bp in length
129356 129255 contig of 8988 bp in length
137104 144376 contig of 7648 bp in length
150760 150659 contig of 6183 bp in length
160623 163263 contig of 5137 bp in length
165295 167704 contig of 4526 bp in length
167805 167704 contig of 2641 bp in length
169851 171185 contig of 1831 bp in length
169851 171185 contig of 1335 bp in length

NOTE: This is a 'working draft' sequence. It currently consists of
18 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved

1 24923 contig of 24923 bp in length
25024 47004 contig of 21981 bp in length
47105 67070 contig of 19966 bp in length
82858 82757 contig of 15887 bp in length
95645 95544 contig of 12687 bp in length
109693 109592 contig of 13948 bp in length
120268 120167 contig of 10475 bp in length
129356 129255 contig of 8988 bp in length
137104 144376 contig of 7648 bp in length
150760 150659 contig of 6183 bp in length
160623 163263 contig of 5137 bp in length
165295 167704 contig of 4526 bp in length
167805 167704 contig of 2641 bp in length
169851 171185 contig of 1831 bp in length
169851 171185 contig of 1335 bp in length

NOTE: This is a 'working draft' sequence. It currently
consists of 18 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 24923: contig of 24923 bp in length
24924 25023: gap of 100 bp
25024 47004: contig of 21981 bp in length
47005 47104: gap of 100 bp
47105 67070: contig of 19966 bp in length
67071 67170: gap of 100 bp
67171 82757: contig of 15587 bp in length
82758 82857: gap of 100 bp
82858 95544: contig of 12687 bp in length
95545 95644: gap of 100 bp
95645 109592: contig of 13948 bp in length
109593 109692: gap of 100 bp
109693 120167: contig of 10475 bp in length
120168 120267: gap of 100 bp
120268 129255: contig of 8988 bp in length
129256 129355: gap of 100 bp
129356 137003: contig of 7648 bp in length
137004 137103: gap of 100 bp
137104 144376: contig of 7273 bp in length
144377 144476: gap of 100 bp
144477 150659: contig of 6183 bp in length
150660 150759: gap of 100 bp
150760 155896: contig of 5137 bp in length
155897 155996: gap of 100 bp
155997 160522: contig of 4526 bp in length
160523 160622: gap of 100 bp
160623 163263: contig of 2641 bp in length
163264 163363: gap of 100 bp
163364 165194: contig of 1831 bp in length
165195 165294: gap of 100 bp
165295 167704: contig of 2410 bp in length
167705 167804: gap of 100 bp
167805 169750: contig of 1946 bp in length
169751 169850: gap of 100 bp
169851 171185: contig of 1335 bp in length.

FEATURES

Source

1. .171185
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q21"
/clone="RP11-873L22"

misc_feature

1. .24923
/note="assembly_fragment"

misc_feature

25024..47004
/note="assembly_fragment"

misc_feature

47105..67070
/note="assembly_fragment"

misc_feature

67171..82757
/note="assembly_fragment"

misc_feature

82858..95544
/note="assembly_fragment"

misc_feature

95645..109592
/note="assembly_fragment"

misc_feature

109693..120167
/note="assembly_fragment"

misc_feature

120268..129255
/note="assembly_fragment"

misc_feature

129356..137003
/note="assembly_fragment clone_end:SP6 vector_side:right"

misc_feature

137104..144376
/note="assembly_fragment"

misc_feature

144477..150659
/note="assembly_fragment"

misc_feature

150760..155896
/note="assembly_fragment"

misc_feature

155997..160522
/note="assembly_fragment"

misc_feature

160623..163263
/note="assembly_fragment"

misc_feature

163364..165194
/note="assembly_fragment clone_end:T7 vector_side:left"

misc_feature

165295..167704


```
repeat_region /map="18"
repeat_region /clone="hrpk_411_h_24"
repeat_region /clone_lib="RPCI-11 human BAC library"
repeat_region 898..1014
repeat_region /rpt_family="L1MD3"
repeat_region complement(1035..1333)
repeat_region /rpt_family="AluSx"
repeat_region 2987..3030
repeat_region /rpt_family="CT-rich"
repeat_region 3087..3291
repeat_region /rpt_family="L1MC4"
repeat_region 4298..4330
repeat_region /rpt_family="(CA)n"
repeat_region 5344..5365
repeat_region /rpt_family="AT-rich"
repeat_region complement(5552..5724)
repeat_region /rpt_family="AluJ"
repeat_region 5731..11847
repeat_region /rpt_family="L1PA2"
repeat_region 11857..11891
repeat_region /rpt_family="(TAAA)n"
repeat_region complement(11893..12002)
repeat_region /rpt_family="AluJ"
repeat_region 11925..11948
repeat_region /note="Single-stranded coverage."
repeat_region 12026..12321
repeat_region /rpt_family="AluSg"
repeat_region complement(12605..12846)
repeat_region /rpt_family="MIR"
repeat_region 13352..13660
repeat_region /rpt_family="AluSc"
repeat_region complement(13741..14128)
repeat_region /rpt_family="MLT1B"
repeat_region 14425..14456
repeat_region /rpt_family="AT-rich"
repeat_region 14457..14978
repeat_region /rpt_family="MLT1F"
repeat_region 15496..15834
repeat_region /rpt_family="THE1B"
repeat_region 16325..16628
repeat_region /rpt_family="AluYb8"
repeat_region 16753..17086
repeat_region /rpt_family="MER7A"
repeat_region 17543..17601
repeat_region /rpt_family="MIR"
repeat_region 18075..18442
repeat_region /rpt_family="MLT1B"
repeat_region complement(18735..18882)
repeat_region /rpt_family="MIR"
repeat_region complement(18948..19083)
repeat_region /rpt_family="L1MB8"
repeat_region complement(19190..19426)
repeat_region /rpt_family="AluJo"
repeat_region 19431..19750
repeat_region /rpt_family="L1MB3"
repeat_region 20434..20555
repeat_region /rpt_family="MLT1A1"
repeat_region 20556..20863
repeat_region /rpt_family="AluSx"
repeat_region 20864..21080
repeat_region /rpt_family="MLT1A1"
repeat_region 21528..21565
repeat_region /rpt_family="AT-rich"
repeat_region 21575..21668
repeat_region /rpt_family="L2"
repeat_region complement(21849..21991)
repeat_region /rpt_family="MER5B"
repeat_region 23147..23168
repeat_region /rpt_family="(TTTG)n"
repeat_region complement(23169..23384)
repeat_region /rpt_family="AluJo"
repeat_region 23385..23419
repeat_region /rpt_family="AT-rich"
```

```
repeat_region complement(23799..23846)
repeat_region /rpt_family="L2"
repeat_region complement(23947..24060)
repeat_region /rpt_family="L2"
repeat_region 24066..24206
repeat_region /rpt_family="BC200"
repeat_region complement(24347..24514)
repeat_region /rpt_family="MER21B"
repeat_region 24782..25203
repeat_region /rpt_family="L1MB2"
repeat_region 25385..26005
repeat_region /rpt_family="L1MB2"
repeat_region 26035..26336
repeat_region /rpt_family="AluSx"
repeat_region 26432..26604
repeat_region /rpt_family="L1ME"
repeat_region 26610..26645
repeat_region /rpt_family="AT-rich"
repeat_region complement(27113..27422)
repeat_region /rpt_family="AluSp"
repeat_region 27865..28190
repeat_region /rpt_family="AluSg"
repeat_region 28250..28272
repeat_region /rpt_family="AT-rich"
repeat_region 28363..28496
repeat_region /rpt_family="HAL1"
repeat_region complement(28497..28796)
repeat_region /rpt_family="AluSg"
repeat_region 28797..28869
repeat_region /rpt_family="HAL1"
repeat_region 28993..29027
repeat_region /rpt_family="AT-rich"
repeat_region 29107..29134
repeat_region /rpt_family="(TATATG)n"
repeat_region 29151..29182
repeat_region /rpt_family="(TG)n"
repeat_region complement(29821..30164)
repeat_region /rpt_family="MER7A"
repeat_region complement(30236..30522)
repeat_region /rpt_family="AluJo"
repeat_region 30576..30723
repeat_region /rpt_family="MIR"
repeat_region 30927..31271
repeat_region /rpt_family="MLT1K"
repeat_region 32382..32437
repeat_region /rpt_family="(CA)n"
repeat_region 32480..32502
repeat_region /rpt_family="(TTTG)n"
repeat_region complement(32503..32751)
repeat_region /rpt_family="AluSx"
repeat_region 34312..34476
repeat_region /rpt_family="MIR"
repeat_region complement(34648..34953)
repeat_region /rpt_family="AluSx"
repeat_region complement(35550..35947)
repeat_region /rpt_family="THE1B"
repeat_region 36048..36070
repeat_region /rpt_family="(TTTAA)n"
repeat_region complement(36072..36366)
repeat_region /rpt_family="AluJo"
repeat_region 36525..36587
repeat_region /rpt_family="AT-rich"
repeat_region 36792..36849
repeat_region /rpt_family="AT-rich"
repeat_region complement(36903..37092)
repeat_region /rpt_family="L1MA7"
repeat_region 37502..37812
```

Query Match

Best Local Similarity 79.0%; Score 22.8; DB 9; Length 179726;

Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 ctcgagccacccataaccctcaatactccagg 34

Db 78718 CTCTAGCCTCCCCATGCCCCCTCATTTCTCCAGTG 78685

- * as soon as it is available and the accession number will
- * be preserved.

804	903:	contig of 803 bp in length
904	1787:	contig of 100 bp
1788	1887:	contig of 884 bp in length
1888	3003:	gap of 100 bp
3004	3103:	contig of 1116 bp in length
3104	3103:	gap of 100 bp
3104	5498:	contig of 2395 bp in length
5499	5598:	contig of 100 bp
5599	7587:	contig of 1989 bp in length
7588	7687:	gap of 100 bp
7688	9739:	contig of 2052 bp in length
9740	9839:	gap of 100 bp
9840	13254:	contig of 3415 bp in length
13255	13354:	gap of 100 bp
13355	17739:	contig of 4385 bp in length
17740	17839:	gap of 100 bp
17840	25230:	contig of 7391 bp in length
25231	25330:	gap of 100 bp
25331	32077:	contig of 6747 bp in length
32078	32177:	gap of 100 bp
32178	48716:	contig of 16539 bp in length
48717	48816:	gap of 100 bp
48817	68998:	contig of 20182 bp in length
68999	69098:	gap of 100 bp
69099	103010:	contig of 33912 bp in length
103011	103110:	gap of 100 bp
103111	182411:	contig of 79301 bp in length

FEATURES	Location/Qualifiers
source	1. 182411

```
/clone lib="RPCT-11 Human Male BAC"
```

misc feature 1 803
/clone_110- KFC1-11 HUMAN MATE BAC

```
misc_feature
T: 003
/note="assembly fragment"
```

```
misc feature 904 1787
/note= assembly_fragment
```

```

misc_retrieve
904: .1/67
/note="assembly fragment"

```

```
misc feature 1998 3003
/note=assembly_fragment"
```

```

MISC_feature
1888. :3003
/note="assembly fragment"

```

```

misc feature      2104  5408
/note="assembly_fragment"

```

```
misc_feature 3104. 5498
```

```

1111 /note="assembly_fragment"
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
19
```

```
misc_feature
5599.7587
```

```

/note="assembly_fragment"

```

```
misc_feature 7688. 9739
```

```
/note="assembly_fragment"
```

misc_feature 9840.13254

Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L12718

Center clone name: 859 C 21

----- Summary statistics

Summary statistics
Sequencing vector: Plasmid: n/a: 100% of reads

sequencing vector: pFastMid; 11/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads

Clientelista: dye-cerminator Big dye; 1008 of
Assembl'y program: Phran: version 0 960731

Assembly program: PHRAP; version 0.960731
Consensus quality: 177348 bases at least 0.99

Consensus quality:	177348	bases at
Consensus quality:	170969	bases at

Consensus quality: 179869 bases at least Q30
Consensus quality: 180657 bases at least Q30

Consensus quality: 180657 bases at least Q20
Insert size: 191111, sum-of-squares

Insert size: 18111; sum-of-contigs: 52 in 320 bases. Quality: coverage: 52 in 320 bases.

Quality coverage: 6.2 in Q20 bases; sum-of-contigs

1
3
r
3
3
3
+
-
E
E
C

NOTE: This is a 'working draft' sequence. It current

consists of 14 contigs. The true order of the pieces

is not known and their order in this sequence record

Query Match

Best Local Similarity 79.48; Pred. No. 33;

Matches	27;	Conservative	0;	Mismatches	7;	Indels	0;	Gaps
<p> </p>								

5

```

Qy 1 ctcagcaccacccataaacctcctcaactccaggg 34
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 122123 CTCAGCCTCCCATGCCCTCATTTCTCCAGTG 122090

RESULT 15
AL159996
LOCUS
DEFINITION
    HUMAN DNA sequence from clone RP11-401F12 on chromosome 9. Contains
    STSs and GSSs, complete sequence.
ACCESSION
    AL159996
VERSION
    AL159996.7
KEYWORDS
    GI:8648450
SOURCE
    human.
ORGANISM
    Homo sapiens
REFERENCE
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS
    1 (bases 1 to 188357)
TITLE
    (bases 1 to 188357)
JOURNAL
    Direct Submission
COMMENT
    Submitted (19-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,
    CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
    requests: clonesrequest@sanger.ac.uk
    On Jun 21, 2000 this sequence version replaced gi:8346239.
    During sequence assembly data is compared from overlapping clones.
    Where differences are found these are annotated as variations
    together with a note of the overlapping clone name. Note that the
    variation annotation may not be found in the sequence submission
    corresponding to the overlapping clone, as we submit sequences with
    only a small overlap as described above.
    This sequence has been finished according to sequence map criteria
    as follows. An attempt is made to resolve all sequencing problems,
    such as compressions and repeats, but not necessarily within known
    annotated human repeat sequence elements (e.g. Alu). Where the
    sequence is ambiguous, there is an annotation using the 'unsure'
    feature key.
    The following abbreviations are used to associate primary accession
    numbers given in the feature table with their source databases:
    Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
    on the WORMPEP database can be found at
    http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence
    was generated from part of bacterial clone contigs of human
    chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
    Group. Further information can be found at
    http://www.sanger.ac.uk/HGP/Chr-9
    RP11-401F12 is from the library RCI-11.2 constructed at the
    Roswell Park Cancer Institute by the group of Pieter de Jong. For
    further details see http://bacpac.med.buffalo.edu/
    VECTOR: pBACe3.6
    This sequence is the entire insert of clone RP11-401F12 The true
    left end of clone RP11-208G24 is at 167549 in this sequence. The
    true right end of clone RP11-395D3 is at 5200 in this sequence.

FEATURES
    source
        1..188357
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="9"
            /clone="RP11-401F12"
            /clone_lib="RCI-11.2"
            410..897
                /note="match: GSS: Em:AQ606829"
                1734..2348
                    /note="match: GSS: Em:AQ823967"
                    1819..2460
                        /note="LMC4 repeat: matches 6462..7077 of consensus"
                        2551..2614
                            /note="2 copies 32 mer 96% conserved"
                            2690..2698
                                /note="TN10 excised: This sequence represents the
                                duplicated flanking sequence of the TN10."
                                3529..3560
                                    /note="8 copies 4 mer gtgt 96% conserved"
                                    3746..3959
                                        repeat_region

```

Job time: 6659 sec

```
repeat_region 30822..31276
/note="MIR repeat: matches 70. .532 of consensus"
repeat_region 31655..32020
/note="THE1B repeat: matches 1. .364 of consensus"
misc_feature 32181..32674
/note="match: GSS: Em:AQ822292"
repeat_region 32279..32403
/note="MIR repeat: matches 57. .192 of consensus"
misc_feature complement(32857..33263)
/note="match: GSS: Em:AQ139531"
repeat_region 33279..33551
/note="AluSg repeat: matches 1. .301 of consensus"
repeat_region 35349..35388
/note="10 copies 4 mer acac 92% conserved"
repeat_region 35353..35388
/note="3 copies 12 mer 94% conserved"
repeat_region 35759..36002
/note="MIR repeat: matches 2. .246 of consensus"
misc_feature 36104..36469
/note="match: GSS: Em:AQ031409"
repeat_region 37497..37545
/note="MIR repeat: matches 81. .127 of consensus"
repeat_region 37664..38248
/note="MIR2D repeat: matches 1. .553 of consensus"
repeat_region 38259..38373
/note="MIR repeat: matches 122. .249 of consensus"
repeat_region 38334..38392
/note="L2 repeat: matches 2692. .2750 of consensus"
misc_feature 38609..38887
/note="match: STS: Em:G13414"
repeat_region 38781..38991
/note="L1R33 repeat: matches 302. .518 of consensus"
repeat_region 39086..39212
/note="L1R33 repeat: matches 29. .165 of consensus"
repeat_region 39249..39483
/note="MER4D repeat: matches 1. .234 of consensus"
repeat_region 39483..40142
/note="MER4D repeat: matches 359. .1017 of consensus"
repeat_region 40325..40415
/note="MER94 repeat: matches 39. .134 of consensus"
repeat_region 41795..41902
/note="L1R33 repeat: matches 370. .472 of consensus"
repeat_region 41903..42192
/note="AluSg repeat: matches 1. .291 of consensus"
repeat_region 42193..42241
/note="L1R33 repeat: matches 472. .521 of consensus"
repeat_region 43494..43613
/note="L2 repeat: matches 2628. .2748 of consensus"
repeat_region 43695..43882
/note="FRAM repeat: matches 0. .164 of consensus"
repeat_region 44019..44082
/note="L2 repeat: matches 2356. .2428 of consensus"
repeat_region 44083..44384
/note="L1P4A repeat: matches 5841. .6144 of consensus"
repeat_region 44385..44529
/note="L2 repeat: matches 2252. .2356 of consensus"
repeat_region 44541..44628
/note="MIR repeat: matches 128. .229 of consensus"
repeat_region 44631..44747
/note="MER33 repeat: matches 203. .324 of consensus"
repeat_region 44779..44985

Query Match 57.0%; Score 22.8; DB 9; Length 188357;
Best Local Similarity 79.4%; Pred.No. 33;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 7 ccaccataaacctcaatactccaggattggg 40
||||||| | ||||| ||||| |||
Db 56526 CCACCCATGCTCTCAAGACTCCAGGGAAGG 56559
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 15:53:27 ; Search time 4272.65 Seconds
(without alignments)
126.357 Million cell updates/sec

Title: US-09-721-543A-12
Perfect score: 40
Sequence: 1 ctgagccaccataaccctcaatactccaggattggg 40

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	24.6	61.5	296	9	AW845395	AW845395 CM4-CT004
C 2	24.6	61.5	486	10	BF740572	BF740572 QV1-HB003
C 3	23.8	59.5	250	10	T08487	T08487 EST06378 In
C 4	23.6	59.0	337	10	T92497	T92497 ye24b09.r1
C 5	23	57.5	69	10	BE878784	BE878784 601493062
C 6	23	57.5	118	9	AW804230	AW804230 PM3-UM008
C 7	23	57.5	135	9	BE005490	BE005490 CM1-BN011
C 8	23	57.5	147	9	BE005492	BE005492 CM1-BN011
C 9	23	57.5	161	10	N73434	N73434 yz31e12.r1
C 10	23	57.5	208	10	BF090424	BF090424 RCL1-NT003
C 11	23	57.5	223	9	BE089733	BE089733 QV0-BT070
C 12	23	57.5	225	10	BF340702	BF340702 602035458
C 13	23	57.5	227	9	BE005486	BE005486 CM1-BN011
C 14	23	57.5	231	9	AA327421	AA327421 EST30756
C 15	23	57.5	237	10	BF037734	BF037734 601462081
C 16	23	57.5	243	10	EG105201	EG105201 602312948
C 17	23	57.5	249	9	AW804233	AW804233 PM3-UM008

C 18	23	57.5	251	10	BE622136	BE622136 601440862
C 19	23	57.5	252	10	BE288188	BE288188 602383704
C 20	23	57.5	252	10	BE869246	BE869246 601445051
C 21	23	57.5	252	10	BE894967	BE894967 601436172
C 22	23	57.5	260	9	AA101534	AA101534 zn79c11.s
C 23	23	57.5	263	9	AA383010	AA383010 EST96488
C 24	23	57.5	265	10	BE878754	BE878754 601493017
C 25	23	57.5	273	10	BF484202	BF484202 QV0-EN005
C 26	23	57.5	276	10	BE617770	BE617770 601441968
C 27	23	57.5	277	10	BE290247	BE290247 602384948
C 28	23	57.5	289	10	T39447	T39447 ya06a08.r1
C 29	23	57.5	290	10	BF740849	BF740849 QV1-HB003
C 30	23	57.5	299	9	BE178852	BE178852 PM1-HT060
C 31	23	57.5	304	9	AA384401	AA384401 EST97914
C 32	23	57.5	305	9	AA853875	AA853875 NHTBCae09
C 33	23	57.5	307	9	AW804241	AW804241 PM3-UM008
C 34	23	57.5	307	10	T39458	T39458 ya06b07.r2
C 35	23	57.5	308	9	AW804238	AW804238 PM3-UM008
C 36	23	57.5	308	9	AA383479	AA383479 EST96859
C 37	23	57.5	313	10	BE936437	BE936437 RCL1-NT003
C 38	23	57.5	316	10	BF342651	BF342651 602013792
C 39	23	57.5	317	10	BG169657	BG169657 602324628
C 40	23	57.5	318	10	BF968657	BF968657 602270903
C 41	23	57.5	321	10	BI015712	BI015712 PM0-ET020
C 42	23	57.5	322	10	BI013197	BI013197 PM0-ET020
C 43	23	57.5	322	10	BI013200	BI013200 PM0-ET020
C 44	23	57.5	324	10	BF884828	BF884828 PM0-ET020
C 45	23	57.5	326	9	AA361993	AA361993 EST71440

ALIGNMENTS

RESULT 1
AW845395/c
LOCUS CM4-CT0045-180200-512-a09 CT0045 Homo sapiens cDNA, mRNA sequence.
DEFINITION AW845395
ACCESSION AW845395.1 GI:7940912
VERSION EST.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 296)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-cm4-CT0045-180200-512-a09&ts=2000-02-18&tf=1)
Seq primer: puc 18 forward
High quality sequence start: 46
High quality sequence stop: 200.
Location/Qualifiers
1. .296
/organism="Homo sapiens"

```

/db_xref="taxon:9606"
/clone_lib="CT0045"
/dev_stage="Adult"
Note="Organ: colon; Vector: puc18; Site 1: SmaI; Site 2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT      70 a      62 c      76 g      87 t      1 others
ORIGIN

Query Match
Best Local Similarity 61.5%; Score 24.6; DB 9; Length 296;
Matches 30; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 tcgagccaccataaccctcaactccaggattggg 40
||| ||||| ||| ||| ||||| ||||| |||||
DB 279 TCCACCACACTTGAAGCCACAATACTGCAGGGATTGG 241

RESULT 2
LOCUS      BF740572      486 bp      mRNA      linear      EST 10-JAN-2001
DEFINITION QV1-HB0031-071200-561-e06 HB0031 Homo sapiens cDNA, mRNA sequence.
ACCESSION  BF740572
VERSION     BF740572.1 GI:12067248
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 486)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl-QV1&t2-QV1-HB0031-
071200-561-e06&t3=2000-12-07&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 8
High quality sequence stop: 485.
Location/Qualifiers
1. 486
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HB0031"
/dev_stage="Adult"
Note="Organ: bocio_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

```

```

BASE COUNT      129 a      129 c      110 g      118 t
ORIGIN

Query Match
Best Local Similarity 61.5%; Score 24.6; DB 10; Length 486;
Matches 30; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 tcgagccaccataaccctcaactccaggattggg 40
||| ||||| ||| ||| ||||| ||||| |||||
DB 167 TCCAGCCACACTTGAAGCCACAATACTGCAGGGATTGG 205

RESULT 3
LOCUS      T08487      250 bp      mRNA      linear      EST 03-AUG-1993
DEFINITION EST06378 Infant Brain, Bento Soares Homo sapiens cDNA clone HIBBE96
ACCESSION  T08487
VERSION     T08487.1 GI:389515
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 250)
Adams,M.D., Soares,M.B., Kerlavage,A.R., Fields,C. and Venter,J.C.
Rapid cDNA sequencing (expressed sequence tags) from a
directionally cloned human infant brain cDNA library
Nature Genet. 4, 373-380 (1993)
94004965
Contact: Adams, MD
The Institute for Genomic Research
932 Clopper Road, Gaithersburg, MD 20878
Tel: 3018699056
Fax: 3018699423
Email: mdadams@tigr.org
Seq primer: M13 Reverse.
Location/Qualifiers
1. 250
/organism="Homo sapiens"
/db_xref="ATCC (inhost):85011"
/db_xref="taxon:9606"
/clone="HIBBE96"
/clone_lib="Infant Brain, Bento Soares"
BASE COUNT      56 a      84 c      72 g      31 t      7 Others
ORIGIN

Query Match
Best Local Similarity 59.5%; Score 23.8; DB 10; Length 250;
Matches 28; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 4 gagccaccataaccctcaactccaggattggg 39
||| ||||| ||||| ||| ||||| |||||
DB 12 GACACCCGCCAAAACCCCACTCTNCAGGGATTGG 47

RESULT 4
LOCUS      T92497/c      337 bp      mRNA      linear      EST 22-MAR-1995
DEFINITION ye24b09.r1 Stratagene lung (#937210) Homo sapiens cDNA clone
ACCESSION  T92497
VERSION     T92497.1 GI:724410
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 337)
Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chisoe,S., Dietrich,N., DuBuque,T., Favellio,A., Gish,W., Hawkins

```


N/3434/C	N73434	161 bp	mRNA	linear	EST 30-JAN-1997
LOCUS					

```

yz31e12.rl Soares_multiple_sclerosis_2NbHSP Homo sapiens cDNA
clone IMAGE:284686 5', mRNA sequence.
N73434
N73434.1 GI:1230719
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 161)
Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chissos,S., Dietrich,N., DuBuque,T., Favello,A., Gish,M., Hawkins
,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
,B., Morris,M., Parsons,J., Prange,C., Thierkin,L., Rohlfing,T.,
Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevasakis,E.,
Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Warra,M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLM: contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1021 Std Error: 0.00
Seq primer: reverse RT
High quality sequence stop: 145.
Location/Qualifiers
1. .161
/organism="Homo sapiens"
/db_xref="GDB:3903214"
/db_xref="taxon:9606"
/clone="IMAGE:284686"
/clone_lib="Soares_multiple_sclerosis_2NbHSP"
/sex="male"
/tissue.type="multiple sclerosis lesions"
/dev_stage="Age 46"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT73D (Pharmacia) with a modified
polylinker V-type: phagemid; Site_1: Not I; Site_2: Eco RI
; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5',
TCGTTACCACTGCAACTGGGAGCGCCGATTTTGTGTTTGTGTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Rento
Soares and M.Fatima Bonaldo. RNA from 4 multiple sclerosis
lesions from one patient was kindly provided by Dr. Kevin
G. Becker (NINDS/NIH)."
BASE COUNT 36 a 45 c 44 g 36 t
ORIGIN
Query Match 57.5%; Score 23; DB 10; Length 161;
Best Local Similarity 74.4%; Pred. No. 1.1e+02;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY 2 tcgagccaccataaccctcaatactccaggatggg 40
||| ||||| ||||| ||||| ||||| ||||| |||||
Db 91 TCCACCCACACTTGACGCCACAAATCGCAGGGATTGGG 53
RESULT 10
BF090424
LOCUS
BF090424
DEFINITION
RC1-NT0033-110900-111-c09 NT0033 Homo sapiens cDNA, mRNA sequence.
ACCESSION
BF090424
VERSION
BF090424.1 GI:10896134

```

KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL
MEDLINE
COMMENT

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A J G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (<http://www.ludwig.org.br/scripts/gethtml2.pl?ti=&t2=RC1-NT0033-110>)
900-111-c09&t3=2000-09-11&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 208.
Location/Qualifiers
1. .208
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NT0033"
/dev_stage="Adult"
/note="Organ: nervous_tumor; Vector: puc18; Site_1: Smal;
Site_2: SmaI; A mini-library was made by cloning products derived from QRESPEC PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 53 a 62 c 51 g 42 t
ORIGIN

Query Match 57.5%; Score 23; DB 10; Length 208;
Best Local Similarity 74.4%; Pred. No. 1.1e+02;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy- 2 tcagaccacccataacctcaatactactcgggattggg 40
|| |||||| | | | | | | | | | | | | | | | |
Db 115 TCCAGCCACACTTGAGCCACAAACTGCGAGGGATTGTG 153

RESULT 11
BE089733/C

LOCUS
DEFINITION
ACCESION
VERSION
KEYWORDS
SOURCE
ORGANISM

BE089733 223 bp mRNA linear EST 12-JUN-2000
QV0-BT0704-120500-224-c08 BT0704 Homo sapiens cdna, mRNA sequence.
BE089733
BE089733.1 GI:8480168
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 223)

REFERENCE
AUTHORS

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-QV0-BF0704-120)
500-324-c086t3-2000-05-12&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 12
High quality sequence stop: 223.

Location/Qualifiers
1..223

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BF0704"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 47 a 52 c 70 g 54 t
ORIGIN

Query Match 57.5%; Score 23; DB 9; Length 223;
Best Local Similarity 74.4%; Pred. No. 1.le+02;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 tcagaccacccataacctcaatctcaggattggg 40
||| |||| | | | | | | | | | | | | | |
Db 86 TCCACCCACACTTGAAGGCCACAAACTGCAGGGATTGGG 48

RESULT 12
BF340702/c

LOCUS BF340702
DEFINITION 602035458f1 NCI_CGAP_Brn64 Homo sapiens cdna clone IMAGE:4183263
5' , mRNA sequence.
BF340702
BF340702.1 GI:11287151
EST.
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 225)

NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: David N. Louis, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: LLAW9499 row: d column: 16
High quality sequence stop: 225.

0;

0000

2

0;

THIS PAGE BLANK (USPTO)


```

; APPLICATION NUMBER:  US 09/315,372
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:  US 08/735,609
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME:  Ingolia, Diane E.
; REGISTRATION NUMBER:  40,027
; REFERENCE/DOCKET NUMBER:  UM-02484
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:  (415) 705-8410
; TELEFAX:  (415) 397-8338
; INFORMATION FOR SEQ ID NO:  9:
; SEQUENCE CHARACTERISTICS:
; LENGTH:  3364 base pairs
; TYPE:  nucleic acid
; STRANDEDNESS:  double
; TOPOLOGY:  linear
; MOLECULE TYPE:  other nucleic acid
; DESCRIPTION:  /desc = "DNA"
US-09-315-372-9

```

Query Match 50.0%; Score 20; DB 3; Length 3364;
Best Local Similarity 72.2%; Pred. No. 15;
Matches 26; Conservative 0; Mismatches 10; Indels

Qy 5 agccacccataaacccctcaatactccaggattgg 40
 |||| ||| |||| ||| ||||| |||
 Db 2326 AGCCTGCCCTCACCCCTGGATCCACAGGACAGGG 2291

```

1  RESULT      8
2  US-09-244-752-9/c
3  : Sequence 9, Application US/09244752
4  : Patent No. 6063622
5  :
6  : GENERAL INFORMATION:
7  :
8  : APPLICANT: Chamberlain, Jeffrey S.
9  : APPLICANT: Amalfitano, Andrea
10 : APPLICANT: Hauser, Michael A.
11 : APPLICANT: Kumar-Singh, Rajendra
12 : APPLICANT: Hartigan-O'Connor, Dennis J.
13 :
14 : TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
15 :
16 : NUMBER OF SEQUENCES: 15
17 :
18 : CORRESPONDENCE ADDRESS:
19 :
20 : ADDRESSEE: Medlen & Carroll, LLP
21 : STREET: 220 Montgomery Street, Suite 2200
22 : CITY: San Francisco
23 : STATE: California
24 :
25 : COUNTRY: United States Of America
26 :
27 : ZIP: 94104
28 :
29 : COMPUTER READABLE FORM:
30 :
31 : MEDIUM TYPE: Floppy disk
32 :
33 : COMPUTER: IBM PC compatible
34 :
35 : OPERATING SYSTEM: PC-DOS/MS-DOS
36 :
37 : SOFTWARE: PatentIn Release #1.0, Version #1.3
38 :
39 : CURRENT APPLICATION DATA:
40 :
41 : APPLICATION NUMBER: US/09/244,752
42 :
43 : FILING DATE:
44 :
45 : CLASSIFICATION:
46 :
47 : PRIOR APPLICATION DATA:
48 :
49 : APPLICATION NUMBER: 09/735,609
50 :
51 : FILING DATE:
52 :
53 : ATTORNEY/AGENT INFORMATION:
54 :
55 : NAME: Ingollia, Diane E.
56 :
57 : REGISTRATION NUMBER: 40,027
58 :
59 : REFERENCE/DOCKET NUMBER: UM-02484
60 :
61 : TELECOMMUNICATION INFORMATION:
62 :
63 : TELEPHONE: (415) 705-8410
64 :
65 : TELEFAX: (415) 397-8338
66 :
67 : INFORMATION FOR SEQ ID NO: 9:
68 :
69 : SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 3364 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
US-09-244-752-9

```

Query Match 50.0%; Score 20; DB 3; Length 3364;
Best Local Similarity 72.2%; Pred. NO. 15;
Matches 26; Conservative 0; Mismatches 10; Indels

Qy 5 agcaccacataaacctcaataactccaggattggg 40
||||| ||||| ||||| ||||| ||||| ||||| |||||

Dd 2326 AGCCTGCCCTCACCCCTGGATCCACGAGACAGG 2291

```

RESULT          9
US-09-245-497-9/c
; Sequence 9, Application US/09245497
; Patent No. 6083750
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; APPLICANT: Amalfitano, Michael A.
; APPLICANT: Hauser, Michael A.
; APPLICANT: Kumar-Singh, Rajendra
; APPLICANT: Hartigan-O'Connor, Dennis J.
; TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States Of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/245,497
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/735,609
; APPLICATION NUMBER: <B> FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: UM-02484
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3364 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
US-09-245-497-9

```

Query Match 50.0%; Score 20; DB 3; Length 3364;
Best Local Similarity 72.2%; pred. No. 15;
Matches 26; Conservative 0; Mismatches 10; Indels

Qy 5 agccaccataaacctcaataactccagggaattggg 40

Db 2326 AGCTGCCCTCACCTGGATCCACGAGGACAGG 2291

RESULT 10

US-09-221-235-10

; Sequence 10, Application US/09221235

; Patent No. 6043040

; GENERAL INFORMATION:

; APPLICANT: Acton, Susan

; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR

; FILE REFERENCE: MNI-050

; CURRENT APPLICATION NUMBER: US/09/221,235

; CURRENT FILING DATE: 1998-12-28

; EARLIER APPLICATION NUMBER: 09/163,115

; EARLIER FILING DATE:

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 10

; LENGTH: 1864

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (275)..(754)

US-09-221-235-10

Query Match

48.0%; Score 19.2; DB 3; Length 1864;

Best Local Similarity 67.5%; Pred. No. 29;

Matches 27; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 ctcgagcccccataacctcaatactccaggattggg 40

||||| ||||| ||||| ||||| |||||

Db 955 ctcagcattcccaagctcttaattctccataaaatggg 994

RESULT 11

US-09-221-928-10

; Sequence 10, Application US/09221928

; Patent No. 6121030

; GENERAL INFORMATION:

; APPLICANT: Acton, Susan

; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR

; FILE REFERENCE: MNI-050

; CURRENT APPLICATION NUMBER: US/09/221,928

; CURRENT FILING DATE: 1998-12-28

; EARLIER APPLICATION NUMBER: 09/163,115

; EARLIER FILING DATE:

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 10

; LENGTH: 1864

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (275)..(754)

US-09-221-928-10

Query Match

48.0%; Score 19.2; DB 3; Length 1864;

Best Local Similarity 67.5%; Pred. No. 29;

Matches 27; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 ctcgagcccccataacctcaatactccaggattggg 40

||||| ||||| ||||| ||||| |||||

Db 955 ctcagcattcccaagctcttaattctccataaaatggg 994

RESULT 12

US-09-221-527-10

; Sequence 10, Application US/09221527

; Patent No. 6146832

; GENERAL INFORMATION:

; APPLICANT: Acton, Susan

; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR

; FILE REFERENCE: MNI-050

; CURRENT APPLICATION NUMBER: US/09/221,527

; CURRENT FILING DATE: 1998-12-28

; EARLIER APPLICATION NUMBER: 09/163,115

; EARLIER FILING DATE:

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 10

; LENGTH: 1864

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (275)..(754)

US-09-221-527-10

; GENERAL INFORMATION:

; APPLICANT: Acton, Susan

; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR

; FILE REFERENCE: MNI-050

; CURRENT APPLICATION NUMBER: US/09/221,527

; CURRENT FILING DATE: 1998-12-28

; EARLIER APPLICATION NUMBER: 09/163,115

; EARLIER FILING DATE:

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 10

; LENGTH: 1864

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (275)..(754)

US-09-221-527-10

Query Match

48.0%; Score 19.2; DB 3; Length 1864;

Best Local Similarity 67.5%; Pred. No. 29;

Matches 27; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 ctcgagcccccataacctcaatactccaggattggg 40

||||| ||||| ||||| ||||| |||||

Db 955 ctcagcattcccaagctcttaattctccataaaatggg 994

RESULT 13

US-09-221-236-10

; Sequence 10, Application US/09221236

; Patent No. 6146841

; GENERAL INFORMATION:

; APPLICANT: Acton, Susan

; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR

; FILE REFERENCE: MNI-050

; CURRENT APPLICATION NUMBER: US/09/221,236

; CURRENT FILING DATE: 1998-12-28

; EARLIER APPLICATION NUMBER: 09/163,115

; EARLIER FILING DATE: 1998-09-29

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 10

; LENGTH: 1864

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (275)..(754)

US-09-221-236-10

Query Match

48.0%; Score 19.2; DB 3; Length 1864;

Best Local Similarity 67.5%; Pred. No. 29;

Matches 27; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 ctcgagcccccataacctcaatactccaggattggg 40

||||| ||||| ||||| ||||| |||||

Db 955 ctcagcattcccaagctcttaattctccataaaatggg 994

RESULT 14

US-09-221-416-10

; Sequence 10, Application US/09221416

; Patent No. 6153417

; GENERAL INFORMATION:

; APPLICANT: Acton, Susan

; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR

; FILE REFERENCE: MNI-050

; CURRENT APPLICATION NUMBER: US/09/221,416

; CURRENT FILING DATE: 1998-12-28

; EARLIER APPLICATION NUMBER: 09/163,115

; EARLIER FILING DATE:

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 10

; LENGTH: 1864

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (275)..(754)

US-09-221-236-10

```
; EARLIER FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1864
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (275)..(754)
US-09-221-416-10
```

```
Query Match      48.0%; Score 19.2; DB 3; Length 1864;
Best Local Similarity 67.5%; Pred. No. 29;
Matches 27; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
```

```
Qy 1 ctccagccaccataaccctcaatactccaggattggg 40
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 955 ctccagcattcccaagctcttaattctccataaaatggg 994
```

```
RESULT 15
US-09-221-245-10
; Sequence 10, Application US/09221245
; Patent No. 6180358
; GENERAL INFORMATION:
; APPLICANT: Acton, Susan
; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
; FILE REFERENCE: MNI-050
; CURRENT APPLICATION NUMBER: US/09/221,245
; CURRENT FILING DATE: 1998-12-28
; EARLIER APPLICATION NUMBER: US 09/163,115
; EARLIER FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1864
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (275)..(754)
US-09-221-245-10
```

```
Query Match      48.0%; Score 19.2; DB 4; Length 1864;
Best Local Similarity 67.5%; Pred. No. 29;
Matches 27; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
```

```
Qy 1 ctccagccaccataaccctcaatactccaggattggg 40
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 955 ctccagcattcccaagctcttaattctccataaaatggg 994
```

Search completed: June 4, 2002, 17:47:09
Job time: 6582 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 15:57:52 ; Search time 521.5 Seconds
(without alignments)
131.690 Million cell updates/sec

Title: US-09-721-543A-12
Perfect score: 40
Sequence: 1 ctcgagccaccataacccctcaatactccaggtggg 40

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	40	100.0	40	22	AAD10599	DNA ligand #5 for
2	23	57.5	2634	19	AAV59067	Bax inhibitor BI-1
3	23	57.5	2922	21	AAF18054	Lung cancer associ
4	21.2	53.0	121	22	ABA71669	Human foetal liver
5	21.2	53.0	121	22	ABA37784	Probe #16250 for g
6	21.2	53.0	121	22	AAK20013	Human brain expres
7	21.2	53.0	121	22	AAK46055	Human bone marrow
8	21.2	53.0	121	22	AAI51965	Probe #20651 used
9	21.2	53.0	473	22	ABA59145	Human foetal liver

C 10	21.2	53.0	473	22	ABA27923	Probe #6388 for ge
C 11	21.2	53.0	473	22	AAK07330	Human brain expres
C 12	21.2	53.0	473	22	AAK33095	Human bone marrow
C 13	21.2	53.0	473	22	AAI38893	Probe #7579 used t
C 14	20.8	52.0	837	20	AAK37523	Human secreted pro
C 15	20.8	52.0	3896	23	ABL06092	Drosophila melanog
C 16	20.6	51.5	366	22	AAI87560	Human polynucleoti
C 17	20.6	51.5	744	22	AAI96641	Human neuroblastom
C 18	20.6	51.5	1907	22	AAI08058	Human extracellular
C 19	20.4	51.0	5592	22	AAF30935	Spodoptera frugipe
C 20	20	50.0	897	23	AAI91288	DNA encoding novel
C 21	20	50.0	1378	23	AAI81354	DNA encoding novel
C 22	20	50.0	3364	19	AAV07266	Muscle creatine ki
C 23	20	50.0	16217	24	ABL32625	Human immune syste
C 24	19.8	49.5	2832	15	AAQ62186	Cyanobacteria repl
C 25	19.8	49.5	4809	15	AAQ62185	pAq1 plasmid fragm
C 26	19.8	49.5	15196	22	AAK73103	Human immune/haema
C 27	19.8	49.5	15196	22	AAK87548	Human immune/haema
C 28	19.8	49.5	32249	22	ABA20005	Human nervous syst
C 29	19.8	49.5	32249	22	ABA07406	Human pancreatic c
C 30	19.8	49.5	32249	22	AAK91137	Human digestive sy
C 31	19.8	49.5	50442	22	AAK73083	Human immune/haema
C 32	19.8	49.5	50442	22	AAK87551	Human immune/haema
C 33	19.6	48.0	511	23	AAI80311	DNA encoding novel
C 34	19.6	49.0	791	22	AAI91505	Human polynucleoti
C 35	19.6	49.0	325791	22	AAI3104	Human Oestrogen re
C 36	19.4	48.5	282	21	AAA43173	Xenopus secreted e
C 37	19.4	48.5	476	22	ABA57621	Human foetal liver
C 38	19.4	48.5	476	22	AAK05674	Human brain expres
C 39	19.4	48.5	476	22	AAK31208	Human bone marrow
C 40	19.4	48.5	476	22	AAI37184	Probe #5870 used t
C 41	19.4	48.5	674	20	AAK37371	Human secreted pro
C 42	19.4	48.5	1554	20	AAK16675	Xenopus WA545 prot
C 43	19.4	48.5	1617	23	AAI85886	DNA encoding novel
C 44	19.4	48.5	12789	24	AAI16876	Human interleukin
C 45	19.4	48.5	349980	22	AAH64966	C glutamicum codin

ALIGNMENTS

RESULT 1
AAD10599
ID AAD10599 standard; DNA; 40 BP.
XX
AC AAD10599;
XX
DT 24-SEP-2001 (first entry)
XX
DE DNA ligand #5 for human cytomegalovirus.
XX
Ligand; human cytomegalovirus; HCMV infection; herpetic ulcer;
KW pneumonia; gastroenteritis; chorioretinitis; antiviral agent;
KW herpesvirus infection; lesion; ds.
XX
OS Human cytomegalovirus.
OS Synthetic.
XX
PN WO200138341-A1.
XX
PD 31-MAY-2001.
XX
PF 21-NOV-2000; 2000WO-US32259.
XX
PR 24-NOV-1999; 99US-0167509.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Liu F, Wang J, Jiang H;
XX
DR WPI; 2001-367658/38.
XX
PT Polynucleotide ligands useful as anti-viral agents for the treatment of

PT herpesvirus infections e.g. primary or chronic infections with human
XX cytomegalovirus -

PS Claim 8; Page 6; 60pp; English.

XX The invention relates to polynucleotide ligands which bind to
CC herpesvirus, particularly human cytomegalovirus (HCMV) to decrease
CC or block HCMV infection in target cells. These ligands are used as
CC anti-viral agents for the treatment of herpesvirus infections (e.g.
CC lesions and herpetic ulcers) and particularly HCMV associated
CC primary or chronic infections such as pneumonia, gastroenteritis
CC and chorioretinitis. The present sequence is DNA ligand which bind
CC to HCMV to decrease or block HCMV infection in target cells.

XX Sequence 40 BP; 10 A; 15 C; 8 G; 7 T; 0 other;

Query Match 100.0%; Score 40; DB 22; Length 40;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctcgagccaccataacccctcaatactccagggtggg 40
|||||
Db 1 ctcgagccaccataacccctcaatactccagggtggg 40

RESULT 2
AAV59067/c
ID AAV59067 standard; cDNA: 2634 BP.

XX AAV59067;

XX 02-FEB-1999 (first entry)

XX Bax inhibitor BI-1 cDNA.

XX Bax inhibitor; BI-1; human; apoptosis; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
FT CDS 73..786
FT /*tag= a

XX WO9840397-A1.

XX 17-SEP-1998.

XX 13-MAR-1998; 98WO-US05015.

XX 14-MAR-1997; 97US-0818514.

XX (BURN-) BURNHAM INST.

XX Reed JC, Xu Q;

XX WPI; 1998-531519/45.

XX P-PSDB; AAW73136.

XX Bax inhibitor proteins, BI-1 and BI-2 - useful e.g. to modulate
XX cellular apoptotic activity or identify agents altering BI-1 or BI-2
XX binding which can modulate apoptotic activity

XX Claim 2; Page 61-63; 80pp; English.

XX This cDNA clone codes for an inhibitor protein, termed BI-1 (see
XX AAW73136), of the pro-apoptotic protein Bax. Nucleic acids encoding
XX BI-1 and BI-2 (see AAV59068) were identified by suppression of
XX Bax-induced death of yeast cells transformed to express human Bax.
XX A human HepG2 cDNA library was used for library screening. The
XX invention provides vectors, optionally expression or viral vectors,
XX containing BI nucleic acids, and host cells containing these
XX vectors. The nucleic acids encoding BI-1/BI-2 can be used to

CC increase expression of these proteins in cells, or antisense
CC molecules prepared from them used to decrease expression. In
CC these ways, cellular apoptotic activity may be modulated (claimed).
CC The nucleic acids and complementary sequences are also useful as
CC probes to detect BI-encoding nucleic acid molecules in samples.

XX Sequence 2634 BP; 632 A; 626 C; 597 G; 779 T; 0 other;

Query Match 57.5%; Score 23; DB 19; Length 2634;
Best Local Similarity 74.4%; Pred. No. 4.1;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ctcgagccaccataacccctcaatactccagggtggg 40
|||||
Db 2056 TCCACCCACTTGAAGCCACAAACTGCAGGGATTGGG 2018

RESULT 3
AAFI8054/c

ID AAFI8054 standard; DNA; 2922 BP.

XX AAFI8054;

XX 14-MAR-2001 (first entry)

XX Lung cancer associated polynucleotide sequence SEQ ID 73.

XX Human; lung cancer associated protein; neuroprotective; cytostatic;
XX cardioactive; immunomodulatory; muscular active; vulnerary;
XX gastrointestinal; nephrotropic; antineoplastic; gynecological;
XX antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
XX proliferative disorder; wound healing; infectious disease; ds.

XX Homo sapiens.

XX WO200055180-A2.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05918.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (ROSE/) ROSEN C A.

XX Ruben SM;

XX WPI; 2000-587514/55.

XX P-PSDB; AAB58178.

XX Lung cancer associated gene sequences, referred to as lung cancer
XX antigens, useful for treatment, prevention, and diagnosis of disorders
XX such as lung cancer -

XX Claim 1; Page 530-551; 1425pp; English.

XX Polynucleotide sequences AAFI7982 - AAFI8424 encode human lung cancer
XX associated proteins represented in AAB58106 - AAB58548. Lung cancer
XX associated proteins and polynucleotide sequences, their agonists, and
XX antagonists may have neuroprotective; cytostatic; cardioactive;
XX immunomodulatory; muscular active general; vulnerary; gastrointestinal
XX general; nephrotropic; antineoplastic; gynecological; or antibacterial
XX activity. The invention also includes antibodies specific for the
XX protein or polynucleotide sequences. The lung cancer associated
XX polynucleotide sequences may be used for detection of lung cancer,
XX chromosome identification, as chromosome markers, and for numerous other
XX diagnostic or research purposes. The proteins may be used to treat
XX disorders such as neural, immune, muscular, reproductive,
XX gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
XX disorders. The proteins may also be used in the treatment of wounds and
XX infectious diseases. Polynucleotide sequences AAFI8425 - AAFI8433 and

Qy 7 ccaccataacctcaatactccagqattqqq 40

AAK20013/c

PA (MOLE-) MOLECULAR DYNAMICS INC.
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX Claim 25; SEQ ID No 20651; 654pp; English.
XX The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX Sequence 121 BP; 46 A; 20 C; 28 G; 27 T; 0 other;
SQ

Query Match 53.0%; Score 21.2; DB 22; Length 121;
Best Local Similarity 76.5%; Pred. NO. 14;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 7 ccacccataacccctcaatactccaggattggg 40
DB 38 CCTCTCTATAATCCTCATTACTCAGGTAATGGG 5

RESULT 9
ABA59145/c
ID ABA59145 standard; DNA; 473 BP.
XX AC ABA59145;
XX DT 01-FEB-2002 (first entry)
XX DE Human foetal liver single exon nucleic acid probe #7450.
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX OS Homo sapiens.
XX PN WO200157277-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00669.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX Claim 1; SEQ ID NO 7450; 639pp + sequence listing; English.
XX The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid

CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 473 BP; 144 A; 91 C; 95 G; 143 T; 0 other;
SQ

Query Match 53.0%; Score 21.2; DB 22; Length 473;
Best Local Similarity 76.5%; Pred. NO. 18;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 7 ccacccataacccctcaatactccaggattggg 40
DB 347 CCTCTCTATAATCCTCATTACTCAGGTAATGGG 314

RESULT 10
ABA27922/c
ID ABA27922 standard; DNA; 473 BP.
XX AC ABA27922;
XX DT 23-JAN-2002 (first entry)
XX DE Probe #6388 for gene expression analysis in human heart cell sample.
XX KW Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX OS Homo sapiens.
XX PN WO200157274-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00666.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488899/53.
XX Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX Claim 1; SEQ ID NO 6388; 530pp; English.
XX The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 473 BP; 144 A; 91 C; 95 G; 143 T; 0 other;
SQ


```
XX 09-AUG-2001.
PD
XX
XX PF 30-JAN-2001; 2001WO-US00663.
XX
XX PR 04-FEB-2000; 2000US-0180312.
XX
XX PR 26-MAY-2000; 2000US-0207456.
XX
XX PR 30-JUN-2000; 2000US-0608408.
XX
XX PR 03-AUG-2000; 2000US-0632366.
XX
XX PR 21-SEP-2000; 2000US-0234687.
XX
XX PR 27-SEP-2000; 2000US-0236359.
XX
XX PR 04-OCT-2000; 2000GB-0024263.
XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX DR WPI; 2001-488897/53.
XX
XX PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
XX PS Claim 25; SEQ ID No 7579; 654pp; English.
XX
XX CC The present invention relates to single exon nucleic acid probes (SENP).
XX The present sequence is one such probe. The probes are useful for
XX producing a microarray for predicting, measuring and displaying gene
XX expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders.
XX
XX SQ Sequence 473 BP; 144 A; 91 C; 95 G; 143 T; 0 other;

Query Match 53.0%; Score 21.2; DB 22; Length 473;
Best Local Similarity 76.5%; Pred. No. 18;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 7 caaccataaccctcaataactccaggattggg 40
DB 347 CCTCTATATATCCCTACTTACTCAGGTAATGGG 314

RESULT 14
AAX37525/c
ID AAX37525 standard; cDNA; 837 BP.
XX
XX AC AAX37525;
XX
XX DT 06-JUL-1999 (first entry)
XX
XX DE Human secreted protein cDNA fragment containing gene 75.
XX
XX KW Human; secreted protein; treatment; prevention; protein therapy; AIDS;
XX gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
XX developmental abnormality; fetal deficiency; blood disorder; leukemia;
XX immune system disease; autoimmune disease; hepatic disease; lymphoma;
XX renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
XX cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
XX pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
XX arthritis; malignancy; digestive; endocrine; infection; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO9918208-A1.
XX
XX PD 15-APR-1999.
XX
XX PF 01-OCT-1998; 98WO-US20775.
XX
XX PR 02-OCT-1997; 97US-0060884.
XX
XX PR 02-OCT-1997; 97US-0060833.
XX
XX PR 02-OCT-1997; 97US-0060836.
XX
XX PR 02-OCT-1997; 97US-0060837.

PR 02-OCT-1997; 97US-0060838.
PR 02-OCT-1997; 97US-0060839.
PR 02-OCT-1997; 97US-0060843.
PR 02-OCT-1997; 97US-0060862.
PR 02-OCT-1997; 97US-0060866.
PR 02-OCT-1997; 97US-0060874.
XX
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Carter KC, Duan DR, Endress GA, Feng P, Ferrie AM;
PI Florence KA, Greene JM, Janat F, Lafleur DW, Ni J;
PI Rosen CA, Ruben SM, Shi Y, Young P, Yu G;
XX
XX DR WPI; 1999-264022/22.
XX
XX DR P-PSDB; AAY07926.
XX
XX PT New isolated human genes and the secreted polypeptides they encode
XX
XX PS Claim 1a; Page 264; 368pp; English.
XX
XX CC This invention describes novel isolated human genes and the secreted
XX proteins they encode. The products of the invention are useful for
XX preventing, treating or ameliorating medical conditions, e.g. by protein
XX or gene therapy. Also pathological conditions can be diagnosed by
XX determining the amount of the new polypeptides in a sample or by
XX determining the presence of mutations in the new polynucleotides.
XX Specific uses are described for each of the 101 polynucleotides, based
XX on which tissues they are most highly expressed in, and include
XX developing products for the diagnosis or treatment of cancer, tumours,
XX neurodegenerative disorders, developmental abnormalities and fetal
XX deficiencies, blood disorders, leukemias, diseases of the immune system,
XX autoimmune diseases, hepatic and renal disease, lymphomas, inflammation,
XX allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate
XX disease, skeletal or cardiac muscle disorders, pulmonary disorders,
XX transplant rejection, disorders involving osteoclasts such as
XX osteoporosis, arthritis or malignancies, digestive/endocrine disorders,
XX infections and AIDS. The human secreted proteins of the invention are
XX represented in AAY07852-Y07993 and the encoding nucleic acids are
XX represented in AAX37451-X37552.
XX
XX SQ Sequence 837 BP; 171 A; 249 C; 187 G; 228 T; 2 other;

Query Match 52.0%; Score 20.8; DB 20; Length 837;
Best Local Similarity 70.0%; Pred. No. 29;
Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 ctgcagccacccataaccctcaataactccaggattggg 40
DB 250 CTCAGGCAGCCCCAGAACACACTCCAGACGCTGGAGATTGGG 211

RESULT 15
ABL06092
ID ABL06092 standard; cDNA; 3896 BP.
XX
XX AC ABL06092;
XX
XX DT 26-MAR-2002 (first entry)
XX
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 12758.
XX
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
XX OS Drosophila melanogaster.
XX
XX PN WO200171042-A2.
XX
XX PD 27-SEP-2001.
XX
XX PF 23-MAR-2001; 2001WO-US09231.
```

PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX
PA (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB61989.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 12758; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3896 BP; 1147 A; 931 C; 883 G; 935 T; 0 other;

Query Match 52.0%; Score 20.8; DB 23; Length 3896;
Best Local Similarity 70.0%; Pred. NO. 38;
Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 1 ctcgagccacccataaccctcaataactccagggtggg 40
||||| ||||| || || |||||
Db 3333 ctcgagccacccacacacagggcacacggtccacggtggg 3372

Search completed: June 4, 2002, 17:56:12
Job time: 7100 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 15:55:32 ; Search time 2137.99 Seconds
(without alignments)
391.518 Million cell updates/sec

Title: US-09-721-543A-12
Perfect score: 40
Sequence: 1 ctcgagcaccataaccctcaatactccaggattggg 40

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.in.*
- 18: em.mu.*
- 19: em.om.*
- 20: em.or.*
- 21: em.ov.*
- 22: em.pat.*
- 23: em.ph.*
- 24: em.pl.*
- 25: em.ro.*
- 26: em.sts.*
- 27: em.un.*
- 28: em.vi.*
- 29: em.htg_hum.*
- 30: em.htg_inv.*
- 31: em.htg_inv.*
- 32: em.htg_inv.*
- 33: em.htg_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description

C	1	23	57.5	1653	9	AK026605	AK026605 Homo sapi
C	2	23	57.5	2600	9	HSTEGT	X75861 H. sapiens T
C	3	23	57.5	2609	9	BC000916	BC000916 Homo sapi
C	4	23	57.5	2634	6	AR058920	AR058920 Sequence
C	5	23	57.5	2634	6	AR058921	AR058921 Sequence
C	6	23	57.5	2634	6	AR112791	AR112791 Sequence
C	7	23	57.5	2634	6	AR112792	AR112792 Sequence
C	8	23	57.5	197189	2	AC019168	AC019168 Homo sapi
C	9	23	57.5	215867	2	AC084037	AC084037 Homo sapi
C	10	23	57.5	225818	2	AC090998	AC090998 Homo sapi
C	11	22.8	57.0	171185	2	AP002424	AP002424 Homo sapi
C	12	22.8	57.0	177097	2	AP001569	AP001569 Homo sapi
C	13	22.8	57.0	179726	9	AC007052	AC007052 Homo sapi
C	14	22.8	57.0	182411	2	AC090408	AC090408 Homo sapi
C	15	22.8	57.0	188357	9	AL159996	AL159996 Human DNA
C	16	22.8	57.0	189329	2	AC090230	AC090230 Homo sapi
C	17	22.8	57.0	196869	2	AC087535	AC087535 Homo sapi
C	18	22.8	57.0	200774	2	AP001592	AP001592 Homo sapi
C	19	22.4	56.0	137845	2	AC010367	AC010367 Homo sapi
C	20	22.4	56.0	151514	9	HS471M13	Z97198 Homo sapi
C	21	22	55.0	452	11	G22123	G22123 human STS W
C	22	22	55.0	34887	3	TBR012139	AJ012199 Trypanoso
C	23	22	55.0	76727	9	HS821D11	AL021453 Human DNA
C	24	22	55.0	112022	9	HS821D11	AJ009611 Homo sapi
C	25	22	55.0	177540	9	AC006533	AC006538 Homo sapi
C	26	21.8	54.5	145607	2	AC105341	AC105341 Homo sapi
C	27	21.8	54.5	152339	2	AC083931	AC083931 Homo sapi
C	28	21.8	54.5	154165	2	AC087620	AC087620 Homo sapi
C	29	21.8	54.5	154599	2	AC016997	AC016997 Homo sapi
C	30	21.8	54.5	157021	2	AC020991	AC020991 Homo sapi
C	31	21.6	54.0	36415	9	AC093645	AC093645 Homo sapi
C	32	21.6	54.0	37470	9	CNS00YVD	AL096810 Homo sapi
C	33	21.6	54.0	104436	9	HS611N7	AL035663 Human DNA
C	34	21.6	54.0	110000	2	LMFLCHR12_4	Continuation (5 of
C	35	21.6	54.0	122952	2	AL513483	AL513483 Homo sapi
C	36	21.6	54.0	173655	2	OSJN00181	AL622983 Oryza sat
C	37	21.6	54.0	210442	2	AL445199	AL445199 Homo sapi
C	38	21.6	54.0	226860	9	AL359920	AL359920 Human DNA
C	39	21.4	53.5	54525	2	AC103846	AC103846 Homo sapi
C	40	21.4	53.5	70993	2	AC102199	AC102199 Mus muscu
C	41	21.4	53.5	161820	2	AC095091	AC095091 Rattus no
C	42	21.4	53.5	217345	2	AC104393	AC104393 Homo sapi
C	43	21.2	53.0	65824	2	AC099899	AC099899 Mus muscu
C	44	21.2	53.0	70594	2	AC015786	AC015786 Homo sapi
C	45	21.2	53.0	116252	9	AL445228	AL445228 Human DNA

ALIGNMENTS

RESULT	1	AK026605	1653 bp	mRNA	linear	PRI 29-SEP-2000
LOCUS	AK026605/c	Homo sapiens	CDNA: FLJ22952	fls, clone	KAT09742.	
DEFINITION	AK026605	AK026605.1	GI:10439494	oligo capping; fls (full insert sequence).		
ACCESSION	AK026605	Homo sapiens	signet-ring cell carcinoma cell_line:KATO III	CDNA to		
VERSION	AK026605	mRNA, clone:lib:KAT	clone:KAT09742.			
KEYWORDS	AK026605	Homo sapiens				
SOURCE	AK026605	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
ORGANISM	AK026605	1 (sites)				
REFERENCE	AK026605	Watanabe, K., Kumagai, A., Itakura, S., Yamazaki, M., Tashiro, H., Ota, T., Suzuki, Y., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.				
AUTHORS	AK026605	NEDO human cDNA sequencing project				
TITLE	AK026605	Unpublished (2000)				
JOURNAL	AK026605	2 (bases 1 to 1653)				
REFERENCE	AK026605	Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T., Shibahara, T., Tanaka, T. and Nakamura, Y.				
AUTHORS	AK026605	Direct Submission				
TITLE	AK026605	Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio				
JOURNAL	AK026605					

[illegible]

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 197189)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 197189)
Waterston,R.H.
Direct Submission
Submitted (30-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Mar 13, 2000 this sequence version replaced gi:7021812.

COMMENT

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
----- Project Information -----
Center project name: H_NH0161E16
----- Summary Statistics -----
Sequencing vector: M13; 92%
Sequencing vector: plasmid; 8%
Chemistry: Dye-primer ET; 92% of reads
Chemistry: Dye-terminator Big Dye; 8% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 180505 bases at least Q40
Consensus quality: 186042 bases at least Q30
Consensus quality: 189445 bases at least Q20
Insert size: 190000; agarose-fp
Insert size: 194889; sum-of-contigs
Quality coverage: 3.69 in Q20 bases; agarose-fp
Quality coverage: 3.65 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1451: contig of 1451 bp in length
* 1452 1551: gap of unknown length
* 1552 2906: contig of 1355 bp in length
* 2907 3006: gap of unknown length
* 3007 4304: contig of 1298 bp in length
* 4305 4404: gap of unknown length
* 4405 5621: contig of 1217 bp in length
* 5622 5721: gap of unknown length
* 5722 8099: contig of 2378 bp in length
* 8100 8199: gap of unknown length
* 8200 10800: contig of 2601 bp in length
* 10801 10900: gap of unknown length
* 10901 14200: contig of 3300 bp in length
* 14201 14300: gap of unknown length
* 14301 17571: contig of 3271 bp in length
* 17572 17671: gap of unknown length
* 17672 21856: contig of 4185 bp in length
* 21857 21956: gap of unknown length
* 21957 25748: contig of 3792 bp in length
* 25749 25848: gap of unknown length
* 25849 28842: contig of 3994 bp in length
* 28843 29943: gap of unknown length
* 29944 35421: contig of 5479 bp in length
* 35422 35521: gap of unknown length
* 35522 40974: contig of 5453 bp in length
* 40975 41075: gap of unknown length
* 41076 48591: contig of 7517 bp in length
* 48592 48692: gap of unknown length
* 48693 58168: contig of 9477 bp in length
* 58169 58269: gap of unknown length
* 58270 68032: contig of 9764 bp in length
* 68033 77463: gap of unknown length
* 77464 9331: contig of 9331 bp in length

* 77464 77563: gap of unknown length
* 77564 93029: contig of 15466 bp in length
* 93030 93129: gap of unknown length
* 93130 107954: contig of 14825 bp in length
* 107955 108054: gap of unknown length
* 108055 121842: contig of 13788 bp in length
* 121843 121943: gap of unknown length
* 121944 137362: contig of 15420 bp in length
* 137363 137462: gap of unknown length
* 137463 155524: contig of 18062 bp in length
* 155525 155624: gap of unknown length
* 155625 173416: contig of 17792 bp in length
* 173417 173517: gap of unknown length
* 173518 197189: contig of 23673 bp in length.
FEATURES
source
Location/Qualifiers
1. 197189
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-161E16"
1. 1451
/note="assembly_name:Contig3"
1552. 2906
/note="assembly_name:Contig4"
3007. 4304
/note="assembly_name:Contig5"
4405. 5621
/note="assembly_name:Contig6"
5722. 8099
/note="assembly_name:Contig7"
8200. 10800
/note="assembly_name:Contig8"
10901. 14200
/note="assembly_name:Contig9"
14301. 17571
/note="assembly_name:Contig10"
17672. 21856
/note="assembly_name:Contig11"
21957. 25748
/note="assembly_name:Contig12"
25849. 29842
/note="assembly_name:Contig13"
29943. 35421
/note="assembly_name:Contig14"
35522. 40974
/note="assembly_name:Contig15"
41075. 48591
/note="assembly_name:Contig16"
48692. 58168
/note="assembly_name:Contig17"
58269. 68032
/note="assembly_name:Contig18"
68133. 77463
/note="assembly_name:Contig19"
77564. 93029
/note="assembly_name:Contig20"
93130. 107954
/note="assembly_name:Contig21"
clone_end:SP6
vector_side:right
108055. 121842
/note="assembly_name:Contig22"
121943. 137362
/note="assembly_name:Contig23"
137463. 155524
/note="assembly_name:Contig24"
155625. 173416
/note="assembly_name:Contig25"
173517. 197189
/note="assembly_name:Contig26"
BASE COUNT 51312 a 47093 c 47437 g 49028 t 2319 others
ORIGIN

Query Match	57.5%	Score 23;	DB 2;	Length 197189;
Best Local Similarity	74.4%	Pred.	No. 27;	
Matches	29;	Conservative	0;	Mismatches 10; Indels 0; Gaps 0;

QY	2	tgcagccaccctcaacccctcatcattcccaggatggg 40
Db	123952	TCCACCCACACTTCGACCCCAAAAAGTGCAGGGATGGG 123990

RESULT	9
AC084037	
LOCUS	215867 bp DNA linear HTG 31-JAN-2002
DEFINITION	Homo sapiens chromosome 12q clone RP11-382C3, WORKING DRAFT SEQUENCE, 7 unordered pieces.
AC084037	
VERSION	GI:1849688
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE	human.
ORGANISM	Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 215867)

AUTHORS
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C., Alsbrooks,S., Benton,J., Bimarge,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Carton,T.F., Carter,M., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Dunthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Galis,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homci,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobsen,B., Jia,X., Johnson,R., Jollivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozados,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkenwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtarian,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,U., Taylor,C., Taylor,T., Teifrod,D., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,B., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 215867)
Worley,K.C.

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Submitted (11-OCT-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jan 31, 2002 this sequence version replaced gi:15809132.
----- Genome Center

```
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: HCHJ
Center clone name: RP11-382G3
----- Summary Statistics -----
Sequencing vector: Plasmid; M77789
Chemistry: Dye-primer Bcdipy: 2% of reads
Chemistry: Dye-terminator Big Dye: 98% of reads
Assembly program: Phrap; version 0.990329first call to findPhrapList

Consensus quality: 223543 bases at least Q40
Consensus quality: 227846 bases at least Q30
Consensus quality: 229658 bases at least Q20
Estimated insert size: 217093; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 96329: contig of 96329 bp in length
* 96330 96429: gap of unknown length
* 96430 176850: contig of 80431 bp in length
* 176861 176960: gap of unknown length
* 176961 189494: contig of 12534 bp in length
* 189495 189594: gap of unknown length
* 189595 202514: contig of 12920 bp in length
* 202515 202614: gap of unknown length
* 202615 208888: contig of 6274 bp in length
* 208889 213216: gap of unknown length
* 208989 213216: contig of 4228 bp in length
* 213217 215867: gap of unknown length
* 215867 215867: contig of 2551 bp in length.
FEATURES             Location/Qualifiers
     source            1..215867
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="12q"
                     /clone="RP11-382G3",
BASE COUNT          59187 a 51141 c 50650 g 54279 t    610 others
ORIGIN

Query Match      57.5%; Score 23; DB 2; Length 215867;
Best Local Similarity 74.4%; Pred.No. 27;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY      2   tcagccaccgccataacccctcaataatcccaggattggg 40
         ||| | | | | | | | | | | | | | | | | | | | | |
Db 142155 TCCACCCACACTTTGAAGCCCAAACTGCAGGGATTGGG 142193

RESULT 10
AC090998
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
```

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE AUTHORS

1 (bases 1 to 225818)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alzuby, R.K., Amara, J.L., Amaratunga, H.C., Are, J.R., Ayele, M., Banks, T., Barabara, J., Benton, J., Blmage, K., Blankenburg, K., Bonnin, D., Bouack, J., Bowie, S., Brleta, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Chen, G., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.C., Dathorne, S.R., David, R., Deavilla, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinch, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, I., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulsegod, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Mettzer, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, R., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Oragunye, N., Ovaledo, R., Pace, A., Payton, B., Peery, J., Perez, Y., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scher, S., Scott, G., Shen, H., Shoostari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczkyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 225818)
Worley, K.C.

Direct Submission
Submitted (23-MAR-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jan 31, 2002 this sequence version replaced qi:15625966.

INTRODUCTION

```

on Jan 31, 2002. This sequence version replaced gi:136235966.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HCM1
Center clone name: Rp11-161E16
----- Summary Statistics
Sequencing vector: Plasmid; M77789
Chemistry: Dye-terminator Big Dye; 47% of reads
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 259159 bases at least Q40
Consensus quality: 268975 bases at least Q30
Consensus quality: 275062 bases at least Q20
Estimated insert size: 250108; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 4.9x in Q20 bases; sum-of-contigs estimation
-----
NOTE: Estimated insert size may differ from sequence length

```

* NOTE: Estimated insert size may differ from sequence length

```
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html)
*
* NOTE: This sequence may represent more than one clone.
* This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
```

*	71791:	contig of 71791 bp in length
*	71792	71891: gap of unknown length
*	71892	123342: contig of 51451 bp in length
*	123343	123344: gap of unknown length
*	123443	144387: contig of 20945 bp in length
*	144388	144487: gap of unknown length
*	144888	160382: contig of 15895 bp in length
*	160383	160482: gap of unknown length
*	160483	171397: contig of 10915 bp in length
*	171398	171497: gap of unknown length
*	171498	178073: contig of 6576 bp in length
*	176074	178173: gap of unknown length
*	176174	184803: contig of 6630 bp in length
*	184803	184903: gap of unknown length
*	184903	191735: contig of 6832 bp in length
*	184904	191835: gap of unknown length
*	191736	196019: contig of 4184 bp in length
*	191836	196019: contig of 4184 bp in length
*	196020	196119: gap of unknown length
*	196120	199454: contig of 3335 bp in length
*	199455	199554: gap of unknown length
*	199555	202369: contig of 3015 bp in length
*	202570	202669: gap of unknown length
*	202670	205014: contig of 2345 bp in length
*	205015	205114: gap of unknown length
*	205115	207645: contig of 2531 bp in length
*	207646	207745: gap of unknown length
*	207746	210468: contig of 2723 bp in length
*	210469	210568: gap of unknown length
*	210569	213268: contig of 2700 bp in length
*	213269	213368: gap of unknown length
*	213369	216260: contig of 2892 bp in length
*	216261	216360: gap of unknown length
*	216361	218775: contig of 2415 bp in length
*	218776	218875: gap of unknown length
*	218876	221207: contig of 2332 bp in length
*	221208	221307: gap of unknown length
*	221308	223378: contig of 2271 bp in length
*	223379	223678: gap of unknown length
*	223679	225818: contig of 2140 bp in length.

	FEATURES	SOURCE
1.	100% Pure Cotton	USA
2.	Machine Washable	India
3.	Soft & Comfortable	Bangladesh
4.	Durable & Long-lasting	Pakistan
5.	Eco-friendly	Vietnam
6.	Available in Multiple Colors	Thailand
7.	Lightweight & Breathable	Malaysia
8.	Easy to Fold & Store	Singapore
9.	Hypoallergenic	Philippines
10.	Non-toxic Dyes	Indonesia

```
1. 223010
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12q"
/clone="RP11-161E16"
```

BASE COUNT	60010 a	53085 c	53539 g	57250 t	1934 others
ORIGIN					

Query Match	57.5%;	Score 23;	DB 2;	Length 235818;
Best Local Similarity	74.4%;	Pred. No. 27;		
Matches 29;	Conservative	0;	Mismatches	10;
			Indels	0;
			Gaps	0;

0v 2 tcgaagccaccataaacctcaatactccaggaattggg 40

Accession	Sequence	Length
Uy	2 ccgagccaccccccataaacctcctcaatactccagggatttggg	40
Db	53174 TCCACCCACACTTGAAGCCACAAAACATGCAGGGATTGGG	53212

RESULT 11

AP002424

AF002424
LOCUS

DEFINITION Homo sapiens chromosome 18 clone RP11-873L22 map 18q21, WORKING DRAFT SEQUENCE, 18 unordered pieces.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AP002424
AP002424.1 GI:8131688
HTG: HTGS_PHASE1; HTGS_DRAFT.
Homo sapiens DNA, clone:RP11-873L22.
Homo sapiens

REFERENCE
AUTHORS

Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;
Mammalia: Eutheria: Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 171185)

JOURNAL
REFERENCE
AUTHORS

Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.,
Homo sapiens 171,185 genomic DNA of 18q21
Published Only in DataBase (2000) In press

TITLE
JOURNAL

2 (bases 1 to 171185)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.,
Direct Submission
Submitted (29-MAY-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@gsc.riken.go.jp,
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)

COMMENT

----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: HumDraft18
Center clone name: RP11-873L22
----- Summary Statistics

Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 156977 bases at least Q40
Consensus quality: 164110 bases at least Q30
Consensus quality: 167421 bases at least Q20
Insert size: 169485; sum-of-contigs
Quality coverage: 4.80x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of
18 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved

1. 24923 contig of 24923 bp in length
25024 25024 contig of 21981 bp in length
47105 47105 contig of 19966 bp in length
67171 67171 contig of 15887 bp in length
82858 82858 contig of 12687 bp in length
95645 95645 contig of 13948 bp in length
109693 120167 contig of 10475 bp in length
120268 129255 contig of 9988 bp in length
129356 137003 contig of 7273 bp in length
137104 144376 contig of 6183 bp in length
144477 150659 contig of 5137 bp in length
150760 155896 contig of 4526 bp in length
155997 160522 contig of 2641 bp in length
160623 163263 contig of 1831 bp in length
163364 165194 contig of 2410 bp in length
165295 167704 contig of 1946 bp in length
167805 169750 contig of 1335 bp in length
169851 171185 contig of 1335 bp in length.

NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 24923: contig of 24923 bp in length
* 24924 25023: gap of 100 bp
* 25024 47004: contig of 21981 bp in length
* 47005 47104: gap of 100 bp
* 47105 67070: contig of 19966 bp in length
* 67071 67170: gap of 100 bp
* 67171 82757: contig of 15887 bp in length
* 82758 82857: gap of 100 bp
* 82858 95544: contig of 12687 bp in length
* 95545 95644: gap of 100 bp
* 95645 109592: contig of 13948 bp in length
* 109593 109692: gap of 100 bp
* 109693 120167: contig of 10475 bp in length
* 120168 120267: gap of 100 bp
* 120268 129255: contig of 9988 bp in length
* 129256 129355: gap of 100 bp
* 129356 137003: contig of 7273 bp in length
* 137004 137103: gap of 100 bp
* 137104 144376: contig of 6183 bp in length
* 144377 144476: gap of 100 bp
* 144477 150659: contig of 5137 bp in length
* 150660 150759: gap of 100 bp
* 150760 155896: contig of 4526 bp in length
* 155897 155996: gap of 100 bp
* 155997 160522: contig of 2641 bp in length
* 160523 160622: gap of 100 bp
* 160623 163263: contig of 1831 bp in length
* 163264 163363: gap of 100 bp
* 163364 165194: contig of 2410 bp in length
* 165195 165294: gap of 100 bp
* 165295 167704: contig of 1946 bp in length
* 167705 167804: gap of 100 bp
* 167805 169750: contig of 1335 bp in length
* 169751 169850: gap of 100 bp
* 169851 171185: contig of 1335 bp in length.

FEATURES

source

1. .171185
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q21"
/clone="RP11-873L22"

misc_feature
1. .24923
/note="assembly_fragment"
25024..47004
/note="assembly_fragment"
47105..67070
/note="assembly_fragment"
67171..82757
/note="assembly_fragment"
82858..95544
/note="assembly_fragment"
95645..109592
/note="assembly_fragment"
109693..120167
/note="assembly_fragment"
120268..129255
/note="assembly_fragment"
129356..137003
/note="assembly_fragment clone_end:SP6 vector_side:right"
137104..144376
/note="assembly_fragment"
144477..150659
/note="assembly_fragment"
150760..155896
/note="assembly_fragment"
155997..160522
/note="assembly_fragment"
160623..163263
/note="assembly_fragment"
163364..165194
/note="assembly_fragment clone_end:T7 vector_side:left"
165295..167704

```

misc_feature      /note="assembly_fragment"
167805..169750
/note="assembly_fragment"
169851..171185
/note="assembly_fragment"
171185..171185
BASE COUNT      52031 a 32086 c 33289 g 52078 t 1701 others
ORIGIN

Query Match      57.0%; Score 22.8; DB 2; Length 171185;
Best Local Similarity 79.4%; Pred. No. 33;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ctcgagccaccccaataaccctcaatactccagg 34
    ||| ||| ||||| ||||| ||||| |||||
Db 93406 CTCAGCTCCCATGCGCCCTCATTTCCAGTG 93439

RESULT 12
LOCUS      AP001569      177097 bp      DNA      linear      HTG 30-MAY-2000
DEFINITION Homo sapiens chromosome 18 clone RP11-859C21 map 18q21, WORKING
DRAFT SEQUENCE, 32 unordered pieces.
ACCESSION      AP001569
VERSION        AP001569.2 GI:8117403
KEYWORDS       HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE        Homo sapiens DNA, clone:RP11-859C21.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 177097)
AUTHORS        Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE          Homo sapiens 177,097 genomic DNA of 18q21
JOURNAL        Published Only in Database (2000) In press
REFERENCE      2 (bases 1 to 177097)
AUTHORS        Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE          Direct Submission
JOURNAL        Submitted (29-MAR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@sc.riken.go.jp,
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)
COMMENT        On May 30, 2000 this sequence version replaced gi:7380904.
-----
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@sc.riken.go.jp
-----
Project Information
Center project name: Humdraft18
Center clone name: RP11-859C21
-----
Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator Ex-amsrham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 160670 bases at least Q40
Consensus quality: 168058 bases at least Q30
Consensus quality: 171631 bases at least Q20
Insert size: 173997; sum-of-contigs
Quality coverage: 4.63x in Q20 bases; sum-of-contigs
-----
NOTE: This is a 'working draft' sequence. It currently consists of
32 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1 23792 contig of 23792 bp in length
23893 45465 contig of 21573 bp in length

```

```

45566 58167 contig of 12602 bp in length
58268 71173 contig of 12906 bp in length
71274 81253 contig of 9980 bp in length
81253 91080 contig of 9727 bp in length
91080 97387 contig of 6207 bp in length
97387 104902 contig of 7415 bp in length
104902 111107 contig of 6105 bp in length
111107 116851 contig of 5644 bp in length
116851 123585 contig of 6634 bp in length
123585 128985 contig of 5300 bp in length
128985 134233 contig of 5148 bp in length
134233 138588 contig of 4255 bp in length
138588 142573 contig of 3885 bp in length
142573 146286 contig of 3613 bp in length
146286 150053 contig of 3667 bp in length
150053 154046 contig of 3893 bp in length
154046 156225 contig of 2079 bp in length
156225 158578 contig of 2253 bp in length
158578 161124 contig of 2446 bp in length
161124 163559 contig of 2335 bp in length
163559 165538 contig of 1879 bp in length
165538 166949 contig of 1311 bp in length
166949 168240 contig of 1191 bp in length
168240 169579 contig of 1239 bp in length
169579 170978 contig of 1299 bp in length
170978 172080 contig of 1002 bp in length
172080 173514 contig of 1334 bp in length
173514 174614 contig of 1000 bp in length
174614 175976 contig of 1262 bp in length
175976 177097 contig of 1021 bp in length
177097 177097 contig of 1021 bp in length

Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 32 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 23792: contig of 23792 bp in length
* 23793 23892: gap of 100 bp
* 23893 45465: contig of 21573 bp in length
* 45466 45565: gap of 100 bp
* 45566 58167: contig of 12602 bp in length
* 58168 58267: gap of 100 bp
* 58268 71173: contig of 12906 bp in length
* 71174 71273: gap of 100 bp
* 71274 81253: contig of 9980 bp in length
* 81254 81353: gap of 100 bp
* 81354 91080: contig of 9727 bp in length
* 91081 91180: gap of 100 bp
* 91181 97387: contig of 6207 bp in length
* 97388 97487: gap of 100 bp
* 97488 104902: contig of 7415 bp in length
* 104903 105002: gap of 100 bp
* 105003 111107: contig of 6105 bp in length
* 111108 111207: gap of 100 bp
* 111208 116851: contig of 5644 bp in length
* 116852 116951: gap of 100 bp
* 116952 123585: contig of 6634 bp in length
* 123586 123685: gap of 100 bp
* 123686 128985: contig of 5300 bp in length
* 128986 129085: gap of 100 bp
* 129086 134233: contig of 5148 bp in length
* 134234 134333: gap of 100 bp
* 134334 138588: contig of 4255 bp in length
* 138589 138688: gap of 100 bp
* 138689 142573: contig of 3885 bp in length
* 142574 142673: gap of 100 bp
* 142674 146286: contig of 3613 bp in length
* 146287 146386: gap of 100 bp
* 146387 150053: contig of 3667 bp in length
* 150054 150153: gap of 100 bp

```



```
/map="18"
/clone="hrpk.411.H_24"
898. .1014
/rpt_family="L1MD3"
complement(1035. .1333)
/rpt_family="AluSx"
2987. .3030
/rpt_family="CT-rich"
3087. .3291
/rpt_family="L1MC4"
4298. .4330
/rpt_family="(CA)n"
5344. .5365
/rpt_family="AT-rich"
complement(5552. .5724)
/rpt_family="AluJ"
5731. .11847
/rpt_family="L1PA2"
11857. .11891
/rpt_family="(TAAA)n"
complement(11893. .12002)
/rpt_family="AluJ"
11925. .11948
/note="Single-stranded coverage."
12026. .12321
/rpt_family="AluSg"
complement(12605. .12846)
/rpt_family="MIR"
13352. .13660
/rpt_family="AluSc"
complement(13741. .14120)
/rpt_family="MLT1B"
14425. .14456
/rpt_family="AT-rich"
14457. .14978
/rpt_family="MLT1P"
15496. .15834
/rpt_family="THE1B"
16325. .16628
/rpt_family="AluYb8"
16753. .17086
/rpt_family="MERT7A"
17543. .17601
/rpt_family="MIR"
18075. .18442
/rpt_family="MLT1B"
complement(18735. .18882)
/rpt_family="MIR"
complement(18948. .19083)
/rpt_family="L1MB8"
complement(19190. .19426)
/rpt_family="AluJo"
19431. .19750
/rpt_family="L1ME3"
20434. .20555
/rpt_family="MLT1A1"
20556. .20863
/rpt_family="AluSx"
20864. .21080
/rpt_family="MLT1A1"
21528. .21565
/rpt_family="AT-rich"
21575. .21668
/rpt_family="L2"
complement(21849. .21991)
/rpt_family="MER5B"
23147. .23168
/rpt_family="(TTTG)n"
complement(23169. .23384)
/rpt_family="AluJo"
23385. .23419
/rpt_family="AT-rich"
```

```
repeat_region complement(23799. .23846)
/rpt_family="L2"
repeat_region complement(23947. .24060)
/rpt_family="L2"
repeat_region 24066. .24206
/rpt_family="BC200"
repeat_region complement(24347. .24514)
/rpt_family="MER21B"
24782. .25203
/rpt_family="L1MB2"
25385. .26005
/rpt_family="L1MB2"
26035. .26336
/rpt_family="AluSx"
26432. .26604
/rpt_family="L1ME"
26610. .26645
/rpt_family="AT-rich"
complement(27113. .27422)
/rpt_family="AluSp"
27865. .28190
/rpt_family="AluSg"
28250. .28272
/rpt_family="AT-rich"
28363. .28496
/rpt_family="HAL1"
complement(28497. .28796)
/rpt_family="AluSg"
28797. .28869
/rpt_family="HAL1"
28993. .29027
/rpt_family="AT-rich"
29107. .29134
/rpt_family="(TATATG)n"
29151. .29182
/rpt_family="(TG)n"
complement(29821. .30164)
/rpt_family="MER7A"
complement(30236. .30522)
/rpt_family="AluJo"
30576. .30723
/rpt_family="MIR"
30927. .31271
/rpt_family="MLT1K"
32382. .32437
/rpt_family="(CA)n"
32480. .32502
/rpt_family="(TTTG)n"
complement(32503. .32751)
/rpt_family="AluSx"
34312. .34476
/rpt_family="MIR"
complement(34648. .34953)
/rpt_family="AluSx"
complement(35350. .35947)
/rpt_family="THE1B"
36048. .36070
/rpt_family="(TTTTA)n"
complement(36072. .36366)
/rpt_family="AluJo"
36525. .36587
/rpt_family="AT-rich"
36792. .36849
/rpt_family="AT-rich"
complement(36903. .37092)
/rpt_family="L1MA"
37502. .37812
```

```
Query Match 57.0%; Score 22.8; DB 9; Length 179726;
Best Local Similarity 79.4%; Pred. No. 33;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Oy 1 ctcgagccacccataacccctcaatactccagg 34
```



```

Qy 1 ctcgagccaccataaacctcaataactccagg 34
   ||| |||| ||||| ||||| ||||| ||||| ||
Db 122123 CTCAGCCTCCCATGCGCCCTCATTCCTCCAGTG 122090

RESULT 15
LOCUS AL159996 188357 bp DNA linear PRI 24-JUL-2000
DEFINITION Human DNA sequence from clone RP11-401F12 on chromosome 9. Contains
ACCESSION STSs and GSSs, complete sequence.
VERSION AL159996
KEYWORDS AL159996.7 GI:8648450
SOURCE HTG.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 188357)
Laird,G.
Direct Submission
Submitted (19-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jun 21, 2000 this sequence version replaced gi:8346239.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C/elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr9
RP11-401F12 is from the library RPCI-11.2 constructed at the
Roswell Park Cancer Institute by the group of Pieter de Jong. For
further details see http://bacpac.med.buffalo.edu/
VECTOR: pBACe3.6
This sequence is the entire insert of clone RP11-401F12 The true
left end of clone RP11-208G24 is at 167549 in this sequence. The
true right end of clone RP11-395D3 is at 5200 in this sequence.
Location/Qualifiers
1..188357
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-401F12"
/clone_lib="RPCI-11.2"
10..897
/misc_feature
/note="match: GSS: Em:AQ0606829"
1734..2348
/misc_feature
/note="match: GSS: Em:AQ823967"
1819..2460
/repeat_region
/note="LIMC4 repeat: matches 6462..7077 of consensus"
2551..2614
/repeat_region
/note="2 copies 32 mer 96% conserved"
2690..2698
/misc_feature
/note="TN10 excised: This sequence represents the
duplicated flanking sequence of the TN10."
3529..3560
/repeat_region
/note="8 copies 4 mer gtgt 96% conserved"
3746..3959
/repeat_region

/feature="L1PB3 repeat: matches 5916..6147 of consensus"
3969..4006
/feature="U2 repeat: matches 1..38 of consensus"
4392..4430
/feature="13 copies 3 mer aac 92% conserved"
5217..5472
/feature="match: GSS: Em:AQ302480"
5260..6024
/feature="match: GSS: Em:AQ901451"
complement(5743..6091)
/feature="match: GSS: Em:AQ132828"
6101..6656
/feature="match: GSS: Em:AQ830958"
7246..8450
/feature="Tiger3b repeat: matches 8..1231 of consensus"
8626..8773
/feature="L2 repeat: matches 2452..2611 of consensus"
10391..10584
/feature="MER5A repeat: matches 1..189 of consensus"
11880..12174
/feature="L2 repeat: matches 1986..2272 of consensus"
complement(12504..12890)
/feature="match: STS: Em:ISA311YCL"
12609..12648
/feature="10 copies 4 mer gtgt 90% conserved"
12610..12647
/feature="19 copies 2 mer tg 92% conserved"
12615..12650
/feature="6 copies 6 mer gtgtgt 91% conserved"
13805..14099
/feature="Alusg repeat: matches 2..295 of consensus"
14101..14388
/feature="Alusq repeat: matches 16..302 of consensus"
15605..15839
/feature="L2 repeat: matches 2357..2612 of consensus"
15947..16013
/feature="L2 repeat: matches 2609..2674 of consensus"
16084..16364
/feature="AluJo repeat: matches 1..288 of consensus"
complement(16254..16783)
/feature="match: GSS: Em:AQ764225"
17436..17528
/feature="MIR repeat: matches 35..123 of consensus"
18177..18298
/feature="FLAM_A repeat: matches 1..122 of consensus"
19380..19675
/feature="Alusx repeat: matches 1..296 of consensus"
complement(19534..19956)
/feature="match: GSS: Em:AQ042003"
19736..19958
/feature="L2 repeat: matches 2477..2745 of consensus"
20126..20237
/feature="L2 repeat: matches 2589..2705 of consensus"
20737..20812
/feature="MADE1 repeat: matches 5..80 of consensus"
21025..21200
/feature="L2 repeat: matches 2184..2355 of consensus"
21443..21646
/feature="MIR repeat: matches 40..246 of consensus"
23474..23793
/feature="MLTIF repeat: matches 197..541 of consensus"
complement(25531..26230)
/feature="match: GSS: Em:AQ787911"
27777..28867
/feature="LIMC3 repeat: matches 6667..7734 of consensus"
28802..28880
/feature="LIMC2 repeat: matches 6247..6326 of consensus"
28924..29133
/feature="MER20 repeat: matches 2..214 of consensus"
29810..29849
/feature="20 copies 2 mer ac 80% conserved"
29868..30157
/feature="Alur repeat: matches 3..290 of consensus"

```

Job time: 6609 sec

```
repeat_region 30822..31276
/note="MLT1H repeat: matches 70. .532 of consensus"
repeat_region 31665..32020
/note="THE1B repeat: matches 1. .364 of consensus"
misc_feature 32181..32674
/note="match: GSS: Em:AQ822292"
repeat_region 32275..32403
/note="MIR repeat: matches 57. .192 of consensus"
misc_feature complement(32857..33263)
/note="match: GSS: Em:AQ139531"
repeat_region 33279..33551
/note="AluSg repeat: matches 1. .301 of consensus"
repeat_region 35349..35388
/note="10 copies 4 mer acac 92% conserved"
repeat_region 35353..35388
/note="3 copies 12 mer 94% conserved"
repeat_region 35759..36002
/note="MIR repeat: matches 2. .246 of consensus"
misc_feature 36104..36469
/note="match: GSS: Em:AQ031409"
repeat_region 37497..37545
/note="MIR repeat: matches 81. .127 of consensus"
repeat_region 37664..38248
/note="MLT2D repeat: matches 1. .553 of consensus"
repeat_region 38259..38373
/note="MIR repeat: matches 122. .249 of consensus"
repeat_region 38334..38392
/note="L2 repeat: matches 2692. .2750 of consensus"
misc_feature 38609..38867
/note="match: STS: Em:G13414"
repeat_region 38781..38991
/note="LTR33 repeat: matches 302. .518 of consensus"
repeat_region 39086..39212
/note="LTR33 repeat: matches 29. .165 of consensus"
repeat_region 39249..39483
/note="MER4D repeat: matches 1. .234 of consensus"
repeat_region 39483..40142
/note="MER4D repeat: matches 359. .1017 of consensus"
repeat_region 40325..40415
/note="MER94 repeat: matches 39. .134 of consensus"
repeat_region 41795..41902
/note="LTR33 repeat: matches 370. .472 of consensus"
repeat_region 41903..42192
/note="AluSg repeat: matches 1. .291 of consensus"
repeat_region 42193..42241
/note="LTR33 repeat: matches 472. .521 of consensus"
repeat_region 43494..43613
/note="L2 repeat: matches 2628. .2748 of consensus"
repeat_region 43695..43862
/note="FRAM repeat: matches 0. .164 of consensus"
repeat_region 44019..44082
/note="L2 repeat: matches 2356. .2428 of consensus"
repeat_region 44083..44384
/note="LIPM4 repeat: matches 5841. .6144 of consensus"
repeat_region 44385..44529
/note="L2 repeat: matches 2252. .2356 of consensus"
repeat_region 44541..44628
/note="MIR repeat: matches 128. .229 of consensus"
repeat_region 44631..44747
/note="MER33 repeat: matches 203. .324 of consensus"
repeat_region 44779..44985
```

```
Query Match 57.0%; Score 22.8; DB 9; Length 188357;
Best Local Similarity 79.4%; Pred. No. 33;
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 7 ccaccccaataacctcaactccagggttggg 40
||||||| | |||| | ||||| |||
Db 56526 CCACCCCATGCTCTCAAGACTCCAGGGAAGG 56559
```

Search completed: June 4, 2002, 17:45:41

THIS PAGE BLANK (USPTO)